2-16-01 PCT

JC09 Rec'd PCT/PTO 14 FEB 2001

FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		ATTORNEY'S DOCKET NUMBER **PF-0579 USN** U.S. APPLICATION NO. (If known, see 37 CFR 1.5) **TO BE ASSIGNED** 09/763233
INTERNATIONAL APPLICATION NO **PCT/US99/19361**	INTERNATIONAL FILING DATE **20 August 1999**	PRIORITY DATE CLAIMED **21 August 1998**

TITLE OF INVENTION
HUMAN RNA-ASSOCIATED PROTEINS

APPLICANT(S) FOR DO/EO/US
INCYTE PHARMACEUTICALS, INC.; HILLMAN, Jennifer L.; YUE, Henry; TANG, Y. Tom; CORLEY, Neil C.; GUEGLER, Karl J.; GORGONE, Gina A.; PATTERSON, Chandra; BAUGHN, Mariah R.; LAL, Preeti; BANDMAN, Olga; REDDY, Roopa; AZIMZAI, Yalda; SHIH, Leo L.; YANG, Junming; LU, Dyung Aina M.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is the **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).
4. ☐ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau)
   b. ☐ has been communicated by the International Bureau
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☒ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included.
13. ☐ A FIRST preliminary amendment.
    ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter
16. ☒ Other items or information.

1) Transmittal Letter (2 pp, in duplicate)
2) Return Postcard
3) Express Mail Label No.: **EL 579 909 774 US**

JC02 Rec'd PCT/PTO 14 FEB 2001

U.S. APPLICATION NO (if known, see 37 CFR 1.5) TO BE ASSIGNED 09/763233	INTERNATIONAL APPLICATION NO · PCT/US99/19361	ATTORNEY'S DOCKET NUMBER PF-0579 USN

17. ☒ The following fees are submitted

BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5):

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1 445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO..... $1000.00

☐International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the FPO or IPO $860.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1 445(a)(2)) paid to USPTO $710.00

☒**International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4).....$690.00**

☐International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4)..... ...$100.00

ENTER APPROPRIATE BASIC FEE AMOUNT =				$690.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e))				$	
CLAIMS	NUMBER FILED		NUMBER EXTRA	RATE	
Total Claims	20	−	0	X **$ 18.00**	$
Independent Claims	2	=	0	X **$ 80.00**	$
MULTIPLE DEPENDENT CLAIM(S) (if applicable)				+ **$270.00**	$
TOTAL OF ABOVE CALCULATIONS =				$690.00	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2				$	
SUBTOTAL =				$690.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1492(f)). +				$	
TOTAL NATIONAL FEE =				$690.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3 28, 3 31). **$40.00** per property +				$	
TOTAL FEES ENCLOSED =				$690.00	
				Amount to be Refunded.	$
				Charged	$

a. ☐ A check in the amount of $_____ to cover the above fees is enclosed.

b. ☒ Please charge my Deposit Account No. 09-0108 in the amount of $ 690.00 to cover the above fees.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 09-0108 . A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

INCYTE GENOMICS, INC
3160 Porter Drive
Palo Alto, CA 94304

SIGNATURE

NAME: Diana Hamlet-Cox

REGISTRATION NUMBER: 33,302

DATE: 14 February 2001

HUMAN RNA-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human RNA-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, and infectious disorders.

BACKGROUND OF THE INVENTION

Ribonucleic acid (RNA) is a linear single-stranded polymer of four nucleotides, ATP, GTP, UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA), the genetic material of the organism. In retroviruses RNA rather than DNA serves as the genetic material. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function. Messenger RNAs (mRNAs) encode polypeptides. Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA translation by recognizing both an mRNA codon and the amino acid that matches that codon.

The unspliced precursors of mature mRNA transcripts are called heterogeneous nuclear RNA (hnRNA) transcripts. hnRNA is generally larger and more unstable than mRNA. Immediately upon its synthesis, hnRNA is assembled into protein-containing complexes called heterogeneous nuclear ribonucleoprotein particles (hnRNPs). (See, for example, Honore, B. et al. (1995) J. Biol. Chem. 270:28780-28789.) hnRNPs associate with small nuclear ribonucleoprotein particles (snRNPs) which are stable RNA-protein complexes that function primarily in splicing introns from hnRNA. Each snRNP contains a single species of RNA and about 10 proteins. The RNA components of snRNPs recognize and base pair with specific sequences of the hnRNA intron. Five different snRNPs associate at the intron of hnRNA to form the spliceosome, a multicomponent RNP complex which catalyzes the removal of introns and the rejoining of exons. Also associated with the snRNPs are various accessory factors that stabilize intron-snRNP interactions. In humans, these factors include spliceosome associated protein 49 (SAP 49) and SAP 145. (Champion-Arnaud, P. and Reed, R. (1994) Genes Dev. 8:1974-1983.)

Proteins are associated with RNA during its transcription from DNA, RNA processing, and translation of mRNA into protein. Proteins are also associated with RNA as it is used for

structural, catalytic, and regulatory purposes. RNA polymerases are proteins that transcribe RNA from a DNA copy. The HIV Tat protein binds specific sites in the viral RNA to prevent premature transcriptional termination. Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The spliceosomal complex is comprised of five small nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base pair with intron consensus sequences. The protein components mediate spliceosome assembly and the splicing reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, L. (1995) Biochemistry W.H. Freeman and Company, New York NY, p. 863).

The process of splicing may involve more than the removal of an intron from an RNA transcript. For example, an RNA transcript may be subject to alternative patterns of splicing, resulting in the generation of different species of mRNA from a single primary transcript. In addition, certain transcripts may be subject to *trans*-splicing, in which an exon from one transcript is joined to an exon of another. Often, specific protein cofactors are required to mediate splicing under these special circumstances. For example, a new splicing factor, PR264, has been implicated in the *trans*-splicing of a thymus-specific c-myb transcript in humans. (Vellard, M. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2511-2515.)

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have been identified that have roles in splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) Clin. Exp. Rheumatol. 16:317-326). Some examples of hnRNPs include the yeast proteins Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) Genes Dev. 12:679-691). HnRNPs have been shown to be
important targets of the autoimmune response in rheumatic diseases (Biamonti, supra).

Nascent tRNA transcripts are spliced by unconventional mechanisms that are distinct from those employed by the spliceosome. In this case, splicing is carried out by specific endonucleases and ligases that recognize secondary structural features of the tRNA. This process contrasts with the spliceosomal reaction, in which specific nucleotide sequences of the intron are recognized. In addition, tRNAs are further processed by removal of 5' sequences and by chemical modification of some of the nucleotide bases. tRNA processing has been extensively studied in yeast, in which tRNA-specific splicing factors have been identified. (See, for example, Shen, W. C. et al. (1993)

J. Biol. Chem. 268:19436-19444.)

Proteins are also a part of the translation machinery of the cell. The eukaryotic ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, the ribosome also contains more than fifty proteins. The ribosomal proteins have a prefix which denotes the subunit to which they belong, either L (large) or S (small). Initiation factors, many of which contain multiple subunits, are proteins which are involved in bringing together an initiator tRNA, the mRNA, and the ribosomal 40S subunit. Eukaryotic initiation factor 2 (eIF2), a guanine nucleotide binding protein, recruits the initiator tRNA to the 40S ribosomal subunit. Only when eIF2 is bound to GTP does it associate with the initiator tRNA. eIF2B, a guanine nucleotide exchange protein, is responsible for converting eIF2 from the GDP-bound inactive form to the GTP-bound active form. Other initiation factors include eIF1A, eIF3, eIF4F (a complex including eIF4E, eIF4A, and eIF4G), and eIF5. The elongation factors EF1α, EF1β γ, and EF2 are involved in elongating the polypeptide chain following initiation, and the release factor eRF carries out termination of translation. (See V. M. Pain (1996) Eur. J. Biochem. 236:747-771.)

Other important RNA-associated enzymes with roles in translation are the aminoacyl-transfer RNA (tRNA) synthetases. Protein biosynthesis depends on each amino acid forming a linkage with the appropriate tRNA. The aminoacyl-tRNA synthetases are responsible for the activation and correct attachment of an amino acid with its cognate tRNA. The 20 aminoacyl-tRNA synthetase enzymes can be divided into two structural classes, and each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding Rossman 'fold'. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet motif, as well as – and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the – and C-terminal regulatory domains (Hartlein, M. and Cusack, S. (1995) J. Mol. Evol. 40:519-530). Autoantibodies against aminoacyl-tRNAs are generated by patients with dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

In many cases, mRNA translation, localization, and stability are controlled by regulatory proteins that bind to the 5' and 3' untranslated (UTR) regions of mRNA. An example of such a protein is Spnr, a mouse spermatid perinuclear RNA-binding protein, which may be involved in RNA transport, translational activation, or localization of RNA to cytoplasmic microtubules

(Schumacher, J. M. et al. (1995) J. Cell Biol. 129:1023-1032). RNA-associated proteins may alter and regulate RNA conformation and secondary structure. These processes are mediated by RNA helicases which utilize energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the DEAD-box family, so named for the conserved B-type ATP-binding motif which is diagnostic of proteins in this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in diverse processes such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability. Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and embryogenesis. All DEAD-box helicases contain several conserved sequence motifs spread out over about 420 amino acids. These motifs include an A-type ATP binding motif, the DEAD-box/B-type ATP-binding motif, a serine/arginine/threonine tripeptide of unknown function, and a C-terminal glycine-rich motif with a possible role in substrate binding and unwinding. In addition, alignment of divergent DEAD-box helicase sequences has shown that 37 amino acid residues are identical among these sequences, suggesting that conservation of these residues is important for helicase function. (Reviewed in Linder, P. et al. (1989) Nature 337:121-122.)

Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors (Godbout, R. et al. (1998) J. Biol. Chem. 273:21161-21168). These observations suggest that DDX1 may promote or enhance tumor progression by altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in tumorigenesis. (Discussed in Godbout, supra.) For example, murine p68 is mutated in ultraviolet light-induced tumors, and human DDX6 is located at a chromosomal breakpoint associated with B-cell lymphoma. Similarly, a chimeric protein comprised of DDX10 and NUP98, a nucleoporin protein, may be involved in the pathogenesis of certain myeloid malignancies.

Ribonucleases (RNases) are RNA-associated enzymes which catalyze the hydrolysis of phosphodiester bonds in RNA chains, thus cleaving the RNA. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid. Such hybrids occur in cells invaded by retroviruses, and RNase H is an important enzyme in the retroviral replication cycle. RNase H domains are often found as a domain associated with reverse transcriptases. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) Nat. Biotechnol. 15:529-536). Regulation of RNase activity is being

investigated as a means to control tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM). (Reviewed in Birney, E. et al. (1993) Nucleic Acids Res. 21:5803-5816.) The RRM is about 80 amino acids in length and forms four β-strands and two α-helices arranged in an α/β sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to snRNP proteins, examples of RNA-binding proteins which contain the above motifs include heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative splicing. Alternative splicing factors include developmentally regulated proteins which have been identified in lower eukaryotes such as Drosophila melanogaster and Caenorhabditis elegans. These proteins play key roles in developmental processes such as pattern formation and sex determination, respectively. (See, for example, Hodgkin, J. et al. (1994) Development 120:3681-3689.) The RRM includes the ribonucleoprotein-1 (RNP-1) RNA binding motif which is found in snRNP proteins, hnRNP proteins, splicing factors, mRNA binding proteins, and transcriptional regulatory proteins. Other hallmarks of RNA binding proteins include regions of repeated arginine and serine residues (RS repeats).

The discovery of new human RNA-associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, and infectious disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human RNA-associated proteins, referred to collectively as "RNAAP" and individually as "RNAAP-1," "RNAAP-2," "RNAAP-3," "RNAAP-4," "RNAAP-5,""RNAAP-6," "RNAAP-7," "RNAAP-8," "RNAAP-9," "RNAAP-10," "RNAAP-11," "RNAAP-12," "RNAAP-13," "RNAAP-14," "RNAAP-15," "RNAAP-16," "RNAAP-17," "RNAAP-18," "RNAAP-19," "RNAAP-20," "RNAAP-21,""RNAAP-22" "RNAAP-23" "RNAAP-24"and RNAAP-25" In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-25 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the

group consisting of SEQ ID NO:1-25 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50 and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ

ID NO:1-25 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-25 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding RNAAP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods and algorithms used for identification of RNAAP.

Table 3 shows the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis, diseases, disorders, or conditions associated with these tissues, and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding RNAAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze RNAAP, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"RNAAP" refers to the amino acid sequences of substantially purified RNAAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which, when bound to RNAAP, increases or prolongs the duration of the effect of RNAAP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of RNAAP.

An "allelic variant" is an alternative form of the gene encoding RNAAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding RNAAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as RNAAP or a polypeptide with at least one functional characteristic of RNAAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding RNAAP, and improper or

unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding RNAAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent RNAAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of RNAAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of RNAAP which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of RNAAP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to RNAAP, decreases the amount or the duration of the effect of the biological or immunological activity of RNAAP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of RNAAP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind RNAAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin,

and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic RNAAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding RNAAP or fragments of RNAAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl),

detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding RNAAP, by northern analysis is indicative of the presence of nucleic acids encoding RNAAP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding RNAAP.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity).

In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Madison WI) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) Parameters for each method may be the default parameters provided by MEGALIGN or may be specified by the user. The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide

sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of RNAAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of RNAAP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:26-50, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:26-50 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:26-50 from related polynucleotide sequences. A fragment of SEQ ID NO:26-50 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:26-50 and the region of SEQ ID NO:26-50 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" and "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding RNAAP, or fragments thereof, or RNAAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters,

chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of RNAAP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to RNAAP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

THE INVENTION

The invention is based on the discovery of new human RNA-associated proteins (RNAAP), the polynucleotides encoding RNAAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, and infectious disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding RNAAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each RNAAP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The clones in column 5 were used to assemble the consensus nucleotide sequence of each RNAAP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows the identity of each protein; and column 7 shows analytical methods used to identify each protein through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding RNAAP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists tissue categories which express RNAAP as a fraction of total tissue categories expressing RNAAP. Column 3 lists diseases, disorders, or conditions associated with those tissues expressing RNAAP. Column 4 lists the vectors used to subclone the cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding RNAAP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The following fragments of the nucleotide sequences encoding RNAAP are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:26-50 and to distinguish between SEQ ID NO:26-50 and related polynucleotide sequences. The useful fragments include the fragment of SEQ ID NO:26 from about nucleotide 586 to about nucleotide

615; the fragment of SEQ ID NO:27 from about nucleotide 399 to about nucleotide 428; the fragment of SEQ ID NO:28 from about nucleotide 234 to about nucleotide 263; the fragment of SEQ ID NO:29 from about nucleotide 40 to about nucleotide 69; and the fragment of SEQ ID NO:30 from about nucleotide 20 to about nucleotide 49, the fragment of SEQ ID NO:31 from about nucleotide 40 to about nucleotide 80, the fragment of SEQ ID NO:32 from about nucleotide 672 to about nucleotide 713, the fragment of SEQ ID NO:33 from about nucleotide 226 to about nucleotide 276, the fragment of SEQ ID NO:34 from about nucleotide 719 to about nucleotide 761, the fragments of SEQ ID NO:35 from about nucleotide 77 to about nucleotide 167 and from about nucleotide 168 to about nucleotide 259, the fragment of SEQ ID NO:36 from about nucleotide 465 to about nucleotide 506, the fragment of SEQ ID NO:37 from about nucleotide 76 to about nucleotide 117, the fragment of SEQ ID NO:38 from about nucleotide 136 to about nucleotide 180, the fragments of SEQ ID NO:39 from about nucleotide 215 to about nucleotide 262, from about nucleotide 683 to about nucleotide 727, and from about nucleotide 1805 to about nucleotide 1852, the fragment of SEQ ID NO:40 from about nucleotide 162 to about nucleotide 206, the fragment of SEQ ID NO:41 from about nucleotide 379 to about nucleotide 423, the fragment of SEQ ID NO:42 from about nucleotide 164 to about nucleotide 208, the fragment of SEQ ID NO:43 from about nucleotide 1 to about nucleotide 42, the fragments of SEQ ID NO:44 from about nucleotide 249 to about 296 and from about nucleotide 816 to about nucleotide 862, the fragment of SEQ ID NO:45 from about nucleotide 196 to about nucleotide 240, the fragment of SEQ ID NO:46 from about nucleotide 1 to about nucleotide 54, the fragment of SEQ ID NO:47 from about nucleotide 463 to about nucleotide 507, the fragments of SEQ ID NO:48 from about nucleotide 551 to about nucleotide 595, from about nucleotide 866 to about nucleotide 910, and from about nucleotide 1406 to about nucleotide 1450, the fragments of SEQ ID NO:49 from about nucleotide 218 to about nucleotide 263, from about nucleotide 758 to about nucleotide 802, and from about nucleotide 1190 to about nucleotide 1234, and the fragment of SEQ ID NO:50 from about nucleotide 11 to about nucleotide 70.

The invention also encompasses RNAAP variants. A preferred RNAAP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the RNAAP amino acid sequence, and which contains at least one functional or structural characteristic of RNAAP.

The invention also encompasses polynucleotides which encode RNAAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:26-50, which encodes RNAAP.

The invention also encompasses a variant of a polynucleotide sequence encoding RNAAP.

In particular, such a variant polynucleotide sequence will have at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding RNAAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:26-50 which has at least about 70%, more preferably at least about 85%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:26-50. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of RNAAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding RNAAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring RNAAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode RNAAP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring RNAAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding RNAAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding RNAAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode RNAAP and RNAAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding RNAAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of

hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:26-50 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 μg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow

fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the Hamilton MICROLAB 2200 (Hamilton, Reno NV), Peltier Thermal Cycler 200 (PTC200; MJ Research, Watertown MA) and the ABI CATALYST 800 (Perkin-Elmer). Sequencing is then carried out using either ABI 373 or 377 DNA sequencing systems (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding RNAAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode RNAAP may be cloned in recombinant DNA molecules that direct expression of RNAAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express RNAAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter RNAAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding RNAAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, RNAAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using

the ABI 431A Peptide Synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of RNAAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active RNAAP, the nucleotide sequences encoding RNAAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding RNAAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding RNAAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding RNAAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding RNAAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding RNAAP. These include, but are not limited to, microorganisms such as

bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding RNAAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding RNAAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding RNAAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of RNAAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of RNAAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of RNAAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) Methods Enzymol. 153:516-54; and Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of RNAAP. Transcription of sequences encoding RNAAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of

Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding RNAAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses RNAAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of RNAAP in cell lines is preferred. For example, sequences encoding RNAAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ or *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers,

e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding RNAAP is inserted within a marker gene sequence, transformed cells containing sequences encoding RNAAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding RNAAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding RNAAP and that express RNAAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of RNAAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on RNAAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul MN, Sect. IV; Coligan, J. E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding RNAAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding RNAAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are

commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding RNAAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode RNAAP may be designed to contain signal sequences which direct secretion of RNAAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Manassas, VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding RNAAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric RNAAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of RNAAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially

available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the RNAAP encoding sequence and the heterologous protein sequence, so that RNAAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled RNAAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of RNAAP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin-Elmer). Various fragments of RNAAP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of RNAAP and RNA-associated proteins. In addition, the expression of RNAAP is closely associated with cancer, fetal development, cell proliferation, inflammation, and immune response. Therefore, RNAAP appears to play a role in cell proliferative, autoimmune/inflammatory, and infectious disorders. In the treatment of disorders associated with increased RNAAP expression or activity, it is desirable to decrease the expression or activity of RNAAP. In the treatment of the above conditions associated with decreased RNAAP expression or activity, it is desirable to increase the expression or activity of RNAAP.

Therefore, in one embodiment, RNAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers

of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and an infectious disorder such as infections by viral agents classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus; infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma; infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, and other fungal agents causing various mycoses; and infections by parasites classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, and cestodes (tapeworm).

In another embodiment, a vector capable of expressing RNAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified RNAAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of RNAAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RNAAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of RNAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of RNAAP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, and infectious disorders described above. In one aspect, an antibody which specifically binds RNAAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express RNAAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding RNAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of RNAAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of RNAAP may be produced using methods which are generally known in the art. In particular, purified RNAAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind RNAAP. Antibodies to RNAAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with RNAAP or with any fragment or

oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to RNAAP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of RNAAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to RNAAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce RNAAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for RNAAP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments

produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between RNAAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering RNAAP epitopes is preferred, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for RNAAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of RNAAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple RNAAP epitopes, represents the average affinity, or avidity, of the antibodies for RNAAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular RNAAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the RNAAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of RNAAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of RNAAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding RNAAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding RNAAP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding RNAAP. Thus, complementary molecules or fragments may be used to modulate RNAAP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding RNAAP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding RNAAP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding RNAAP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding RNAAP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding RNAAP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the

ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding RNAAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding RNAAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits,

monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of RNAAP, antibodies to RNAAP, and mimetics, agonists, antagonists, or inhibitors of RNAAP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated

sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of

RNAAP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example RNAAP or fragments thereof, antibodies of RNAAP, and agonists, antagonists or inhibitors of RNAAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their

inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind RNAAP may be used for the diagnosis of disorders characterized by expression of RNAAP, or in assays to monitor patients being treated with RNAAP or agonists, antagonists, or inhibitors of RNAAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for RNAAP include methods which utilize the antibody and a label to detect RNAAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring RNAAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of RNAAP expression. Normal or standard values for RNAAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to RNAAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of RNAAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding RNAAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of RNAAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of RNAAP, and to monitor regulation of RNAAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding RNAAP or closely related molecules may be used to identify nucleic acid sequences which encode RNAAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies

only naturally occurring sequences encoding RNAAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the RNAAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:26-50 or from genomic sequences including promoters, enhancers, and introns of the RNAAP gene.

Means for producing specific hybridization probes for DNAs encoding RNAAP include the cloning of polynucleotide sequences encoding RNAAP or RNAAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding RNAAP may be used for the diagnosis of disorders associated with expression of RNAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,

thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and an infectious disorder such as infections by viral agents classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and togavirus; infections by bacterial agents classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, and campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, and mycoplasma; infections by fungal agents classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, and other fungal agents causing various mycoses; and infections by parasites classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematodes such as trichinella, intestinal nematodes such as ascaris, lymphatic filarial nematodes, trematodes such as schistosoma, and cestodes (tapeworm). The polynucleotide sequences encoding RNAAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered RNAAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding RNAAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding RNAAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding RNAAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of RNAAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a

sequence, or a fragment thereof, encoding RNAAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding RNAAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding RNAAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding RNAAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of RNAAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously

and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding RNAAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding RNAAP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide

sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, RNAAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between RNAAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with RNAAP, or fragments thereof, and washed. Bound RNAAP is then detected by methods well known in the art. Purified RNAAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding RNAAP specifically compete with a test compound for binding RNAAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with RNAAP.

In additional embodiments, the nucleotide sequences which encode RNAAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/115,639 and U.S. Ser No. 60/097,550, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were

homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-BLUE, XL1-BLUEMRF, or SOLR from Stratagene or DH5α, DH10B, or ELECTROMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision, using the UNIZAP vector system (Stratagene) or cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in

a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a Fluoroskan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing systems (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian,

vertebrate, and eukaryote databases, and BLOCKS to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Str. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:26-50. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding RNAAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular,

dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Extension of RNAAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:26-50 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were

successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:26-50 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:26-50 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following

endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba1, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the RNAAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring RNAAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of RNAAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit

translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the RNAAP-encoding transcript.

IX. Expression of RNAAP

Expression and purification of RNAAP is achieved using bacterial or virus-based expression systems. For expression of RNAAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express RNAAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of RNAAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding RNAAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, RNAAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from RNAAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified RNAAP obtained by these methods can be used directly in the following activity assay.

X. Demonstration of RNAAP Activity

RNAAP activity is demonstrated by the formation of an RNAAP-RNA complex as

detected by a polyacrylamide gel mobility-shift assay. In preparation for this assay, RNAAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing RNAAP cDNA. The cells are incubated for 48-72 hours after transformation under conditions which allow expression and accumulation of RNAAP. Extracts containing solubilized proteins can be prepared from cells expressing RNAAP by methods well known in the art. Portions of the extract containing RNAAP are added to [^{32}P]-labeled RNA. Radioactive RNA can be synthesized in vitro by techniques well known in the art. The mixtures are incubated at 25°C in the presence of RNase inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a high molecular weight band on the autoradiogram indicates the formation of a complex between RNAAP and the radioactive transcript. A band of significantly lower molecular weight will be present in samples prepared using control extracts prepared from untransformed cells. The amount of RNAAP-RNA complex can be quantified using phospho-image analysis and is proportional to the activity of RNAAP.

Alternatively, RNAAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled RNAAP, washed, and any wells with labeled RNAAP complex are assayed. Data obtained using different concentrations of RNAAP are used to calculate values for the number, affinity, and association of RNAAP with the candidate molecules.

XI. Functional Assays

RNAAP function is assessed by expressing the sequences encoding RNAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify

transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of RNAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding RNAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding RNAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of RNAAP Specific Antibodies

RNAAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the RNAAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring RNAAP Using Specific Antibodies

Naturally occurring or recombinant RNAAP is substantially purified by immunoaffinity chromatography using antibodies specific for RNAAP. An immunoaffinity column is constructed by covalently coupling anti-RNAAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing RNAAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of RNAAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/RNAAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and RNAAP is collected.

XIV. Identification of Molecules Which Interact with RNAAP

RNAAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled RNAAP, washed, and any wells with labeled RNAAP complex are assayed. Data obtained using different concentrations of RNAAP are used to calculate values for the number, affinity, and association of RNAAP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	26	399781	PITUNOT02	399781H1 (PITUNOT02), 399781X12 (PITUNOT02), 405935R1 (EOSIHET02), 790764R1 (PROSTUT03), 792124R1 (PROSTUT03), 1271965F6 (TESTTUT02)
2	27	1806542	SINTNOT13	393728R6 (TMLR2DT01), 394324R1 (TMLR2DT01), 659218H1 (BRAINOT03), 1679393T7 (STOMFET01), 1799828F6 (COLNNOT27), 1806542F6 (SINTNOT13), 1806542H1 (SINTNOT13), 2440318H1 (EOSITXT01), 2691085F6 (LUNGNOT23), 2830828F6 (TLYMNOT03), 2881875F6 (UTRSTUT05), 4551743H1 (HELAUNT01), SBDA06024F1
3	28	2263514	UTRSNOT02	097587R1 (PITUNOR01), 1525955F1 (UCMCL5T01), 1733577F6 (BRSTTUT08), 1833563R6 (BRAINON01), 2123362F6 (BRSTNOT07), 2214468F6 (SINTFET03), 2263514H1 (UTRSNOT02)
4	29	2738270	OVARNOT09	1860901F6 (PROSNOT19), 2078366F6 (ISLTNOT01), 2350279T6 (COLSUCT01), 2738270H1 (OVARNOT09), SAEA01159F1
5	30	2824412	ADRETUT06	151898H1 (FIBRAGT01), 151898R6 (FIBRAGT01), 744489R1 (BRAITUT01), 1297568F6 (BRSTNOT07), 1321930F1 (BLADNOT04), 1401380F6 (BRAITUT08), 1417662T1 (KIDNNOT09), 1492985T1 (PROSNON01), 1982147T6 (LUNGTUT03), 2495243H1 (ADRETUT05), 2824412H1 (ADRETUT06)
6	31	002690	HMC1NOT01	002690H1 (HMC1NOT01), 116043F1 (KIDNNOT01), 3181573H1 (TLYJNOT01)

Table 1 (Cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
7	32	041108	TBLYNOT01	041108H1 (TBLYNOT01), 726323R1 (SYNOOAT01), 1445830T1 (PLACNOT02), 1466383F1 (PANCTUT02), 1794401R6 (PROSTUT05), 2817055H1 (BRSTNOT14)
8	33	869138	LUNGAST01	869138H1 (LUNGAST01), 1803154F6 (SINTNOT13), 2080849F6 (UTRSNOT08), 2744328T6 (BRSTTUT14), SAEC11279F1
9	34	934406	CERVNOT01	399781X12 (PITUNOT02), 469628R1 (LATRNOT01),SADA00194R1, SADC12152F1
10	35	1315083	BLADTUT02	896576H1 and 897395R1 (BRSTNOT05), 1315083F1, 1315083H1, and 1315083T6 (BLADTUT02), 1530042R1 (PANCNOT04), 1806531F6 (SINTNOT13), 1810232F6, 1810232X13C1, and 1810232X15C1 (PROSTUT12), 2109147H1 (BRAITUT03), 2885716H1 (SINJNOT02), 3071661H1 (UTRSNOR01), 3217049H1 (TESTNOT07), 3522844H1 (ESOGTUN01)
11	36	1444908	THYRNOT03	211007T6 (SPLNNOT02), 290135H1 (TMLR3DT01), 948920R1 (PANCNOT05), 1444908F1 and 1444908H1 (THYRNOT03), 2640567F6 and 2640567T6 (LUNGTUT08), 4558043H1 (KERAUNT01), 5025039H1 (OVARNON03)
12	37	1557481	BLADTUT04	674210X313V1 and 674210X315V1 (CRBLNOT01), 729909X305D4 (LUNGNOT03), 1365495H1 (SCORNON02), 1524311T1 (UCMCL5T01), 1557481H1 and 1557481T6 (BLADTUT04), 2123062H1 (BRSTNOT07), 2619546R6 (KERANOT02), 3804464H1 (BLADTUT03), 5117228H1 (SMCBUNT01)

Table 1 (Cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
13	38	1747456	STOMTUT02	1737462F6 and 1737462T6 (COLNNOT22), 1747456H1 (STOMTUT02), 2500482H1 (ADRETUT05)
14	39	1748626	STOMTUT02	590849F1 (UTRSNOT01), 874611R1 (LUNGAST01), 1493762R6 (PROSNON01), 1748626H1 (STOMTUT02), 2181685H1 (SININOT01), 2344343H1 (TESTTUT02), 3116348F6 (LUNGTUT13), 3324014H1 (PTHYNOT03), 3661111H1 (ENDPNOT02), 3732550H1 (SMCCNOS01), 4850438H1 (TESTNOT10)
15	40	1879135	LEUKNOT03	1879135F6 and 1879135H1 (LEUKNOT03), 4951531H2 (ENDVUNT01)
16	41	2073417	ISLTNOT01	1338939F6 (COLNTUT03), 2023389X16R1 (CONNNOT01), 2073417H1 (ISLTNOT01), 3039834H1 (BRSTNOT16), SAEA02062F1
17	42	2129080	KIDNNOT05	276066H1 (TESTNOT03), 1834317R6 (BRAINON01), 1932554H1 (COLNNOT16), 2129080H1 and 2129080R6 (KIDNNOT05), 2159002F6 (ENDCNOT02), 2227977F6 (SEMVNOT01), 2232155T6 (PROSNOT16), 2778314H1 (OVARTUT03), 3090406F7 (BRSTNOT19), 4306306H1 (BLADDIT01)

Table 1 (Cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
18	43	2472867	THP1NOT03	791138F1 and 791138R1 (PROSTUT03), 2472867H1 (THP1NOT03)
19	44	2764755	BRSTNOT12	1234994H1 (LUNGFET03), 1834126R6 (BRAINON01), 1868062H1 (SKINBIT01), 2686143T6 (LUNGNOT23), 2764755H1 (BRSTNOT12)
20	45	2875939	THYRNOT10	1261589R1 (SYNORAT05), 2875939H1 and 2875939X331U1 (THYRNOT10)
21	46	3591363	293TF5T01	024088R6 (ADENINB01), 1732992F6 (BRSTTUT08), 3591363H1 (293TF5T01)
22	47	3702292	PENCNOT07	795119F1 and 795119R1 (OVARNOT03), 1639330F6 (UTRSNOT06), 3110107H1 (BRSTTUT15), 3702292H1 (PENCNOT07)
23	48	3778908	BRSTNOT27	002356F1 (U937NOT01), 039273T6 (HUVENOB01), 821404R6 (KERANOT02), 1310730T6 (COLNFET02), 1466025F1 (PANCTUT02), 3778908H1 (BRSTNOT27), 4665703H1 (MEGBUNT01), SBLA01565F1, SBLA02234F1
24	49	4163642	BRSTNOT32	637915R1 (BRSTNOT03), 1505472F1 (BRAITUT07), 1728944X12C1 and 1728944X15C1 (BRSTTUT08), 1737432H1 (COLNNOT22), 2721874T6 (LUNGTUT10), 3989458R6 (LUNGNON03), 4163642H1 (BRSTNOT32)
25	50	4906154	TLYMNOT08	515395F1 (MMLR1DT01), 4906154H2 (TLYMNOT08)

Table 2

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
1	216	S129 T21 S108 T161 T178 T47 S107 S143 T150 S185 Y116 Y138	N9	L12-I83 (RNP-1/RRM RNA binding motif)	PR 264 splicing factor	MOTIFS BLOCKS PROFILESCAN PFAM BLAST
2	962	T169 T384 T46 S130 T161 T174 T197 S21 S577 S588 S646 S725 S426 T440 S537 T661 S666 T670 T750 T903 Y389 Y918			pre-tRNA splicing factor	BLAST
3	285	S28 S46 T182 T17 S45	N246	V136-L204 (RNP-1/RRM RNA binding motif)	hnRNP H protein	PFAM BLOCKS BLAST
4	267	S256 S116 S137 S164 S115 S197 S264 Y238	N124 N183	L12-I82 (RNP-1/RRM RNA binding motif)	SAP49 spliceosomal protein	BLOCKS PFAM BLAST
5	369	S349 S16 T32 S44 S56 S58 S150 T233 S258 T296 T46 S108 S334			Spnr: spermatid perinuclear RNA binding protein	BLAST

Table 2 (Cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
6	175	S140 T81 Y55	N112 N164	Ribosomal protein L17: L28 through K126	ribosomal protein L17	PFAM BLOCKS BLAST
7	311	S110 S23 T116 T222 S229 T57 S68 T133 S140 T179 T234			ribosomal protein L4	BLAST
8	330	S30 S198 S273 S288 T303 S9 T55 S65 S120 S153 S188 S244 S279 S294 S2 T90 S193 S198 S214 S222 S244 S259 S262 S268 S326 S327 Y102 Y124		ATP/GTP-binding site motif A (P-loop): G48 through T55 Zinc finger: S9 through T40 S65 through A94	RNA-associated protein	PFAM MOTIFS BLAST
9	183	S129 T21 S108 S158 S160 T47 S107 S143 T150 T179 S180 Y116 Y138	N9 N166	RNA-recognition motif: L12 through I83 RNP-1: F30 through I81	RNA-binding protein	MOTIFS ProfileScan PFAM BLOCKS BLAST

Table 2 (Cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
10	670	T360 T466 S663 T34 T94 S103 S136 T143 S549 S563 S578 S50 S107 T125 T143 T190 S209 S227 S289 T356 T368 S378 S409 S437 T516 S594 T646 S659 Y447 Y573	N32 N201 N250 N290 N394	ATP/GTP-binding site motif A (P-loop): A223 through T230 DEAD/DEAH box helicase: E196 through K421 Helicases conserved C-terminal domain: K439 through L520	RNA helicase	MOTIFS ProfileScan PFAM BLOCKS BLAST
11	452	T55 T397 S89 T163 S223 S369	N218		initiation factor eIF-2B gamma subunit	BLAST
12	748	T51 S425 T58 T96 T144 T175 S181 S316 T416 S437 T513 S693 S699 S137 T272 S278 S312 S410 T416 S425 S437 T695 Y260	N94 N141 N265 N744	Aminoacyl transfer RNA synthetases class I: S63 through A68 C405 through N413	cysteinyl-tRNA synthetase	BLOCKS BLAST

Table 2 (Cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
13	328	S12 S16 T32 S36 T68 T121 T57 T58 T109 T147	N4 N14	RNP-1: V123 through E169 P220 through R269 RNA-recognition motif: M23 through P93 I112 through C182	RNA-binding protein	MOTIFS BLOCKS PFAM ProfileScan BLAST
14	563	S235 S271 S20 S22 S212 T217 S282 S308 T364 S71 T90 S116 S212 S230 S242 S274 S297 S301 S308 S357 T378 T453 Y170	N101 N312 N383	RNA-recognition motif: L12 through I82 L93 through V160	RNA-binding protein	PFAM BLAST BLOCKS
15	153	T59 S5 T43 S112 S113 S114 S145 T82		Zinc finger: Y26 through T43	hnRNP-associated protein	PFAM BLAST
16	286	S22 T42 T63 S96		RNase H: G137 through K282 Signal Peptide: M1 through C18	RNase H	PFAM BLAST SPScan

Table 2 (Cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
17	537	T149 T6 S30 S56 S81 S118 S421 T41 S118 T144 S145 T231 S280 S421 S488 T526 Y254 Y312	N141 N249 N343 N520	RNA-recognition motif: V73 through I133 L166 through L232 V332 through V399	hnRNP protein	PFAM BLAST
18	163	Y81	N87	Ribosomal protein L24e: M1 through R75	ribosomal protein	PFAM BLAST BLOCKS
19	178	T120 T159 T171 S36		Ribosomal protein L13: A15 through L145	ribosomal protein L13	PFAM BLOCKS BLAST
20	140	S103 T21 R138	N101 N126		RNase P protein component	BLAST
21	209	T113 T163 S40	N144	Ribosomal protein L21p: L99 through L197 Signal Peptide: M1 through S22	ribosomal protein L21	PFAM BLAST SPScan
22	162	T65 T35 T45 T51 S155	N93	Ribosomal protein L23: Y74 through L153	ribosomal protein L23a	ProfileScan PFAM BLAST BLOCKS

Table 2 (Cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
23	623	T580 T29 S55 T348 T101 S115 S230 S308 S524 S531 S575 Y133 Y485	N9 N90	RNA-recognition motif: I164 through V236 L245 through V313 L340 through I403	RNA-binding protein	MOTIFS PFAM BLAST
24	786	S183 T349 T255 T283 T335 T336 S352 S374 S447 S540 S598 T68 S126 S211 S223 S434 S475 S567 S642 T682 T702 T776 Y101	N60 N432 N672	ATP/GTP-binding site motif A (P-loop): G207 through T214 Helicases conserved C-terminal domain: Y379 through S475 Signal Peptide: M1 through G54	RNA helicase	MOTIFS PFAM SPScan BLAST
25	260	S44 S68 T152 S192 Y49		Ribosomal protein S5: V73 through A210 Signal Peptide: M1 through A22	ribosomal protein S5	ProfileScan PFAM BLOCKS SPScan BLAST MOTIFS

Table 3

Seq ID NO:	Tissue Expression (Fraction of Total)	Disease Class (Fraction of Total)	Vector
26	Reproductive (0.309) Nervous (0.191) Cardiovascular (0.106)	Cancer/Cell Proliferation (0.660) Inflammation (0.202) Trauma (0.085)	PSPORT1
27	Reproductive (0.233) Nervous (0.189) Hematopoietic/Immune (0.144)	Cancer/Cell Proliferation (0.756) Inflammation (0.222) Trauma (0.089)	pINCY
28	Reproductive (0.273) Nervous (0.164) Hematopoietic/Immune (0.131)	Cancer/Cell Proliferation (0.694) Inflammation (0.235) Trauma (0.082)	PSPORT1
29	Reproductive (0.326) Gastrointestinal (0.130) Musculoskeletal (0.109)	Cancer/Cell Proliferation (0.804) Inflammation (0.217) Trauma (0.022)	pINCY
30	Nervous (0.245) Reproductive (0.235) Gastrointestinal (0.112)	Cancer/Cell Proliferation (0.663) Inflammation (0.245) Trauma (0.133)	pINCY
31	Reproductive (0.255) Nervous (0.224) Cardiovascular (0.102)	Cell Proliferation (0.673) Inflammation and Immune Response (0.357)	PBLUESCRIPT
32	Reproductive (0.222) Nervous (0.141) Cardiovascular (0.131) Gastrointestinal (0.131)	Cell Proliferation (0.707) Inflammation and Immune Response (0.323)	PBLUESCRIPT
33	Hematopoietic/Immune (0.210) Gastrointestinal (0.161) Nervous (0.161) Reproductive (0.161)	Cell Proliferation (0.645) Inflammation and Immune Response (0.468)	PSPORT1
34	Reproductive (0.301) Nervous (0.219) Cardiovascular (0.123) Hematopoietic/Immune (0.123)	Cell Proliferation (0.644) Inflammation and Immune Response (0.288)	PSPORT1

Table 3 (Cont.)

Seq ID NO:	Tissue Expression (Fraction of Total)	Disease Class (Fraction of Total)	Vector
35	Reproductive (0.220) Gastrointestinal (0.171) Hematopoietic/Immune (0.171) Nervous (0.159)	Cell Proliferation (0.610) Inflammation and Immune Response (0.463)	pINCY
36	Developmental (0.156) Hematopoietic/Immune (0.156) Reproductive (0.156) Nervous (0.111)	Cell Proliferation (0.711) Inflammation and Immune Response (0.356)	pINCY
37	Reproductive (0.230) Nervous (0.170) Gastrointestinal (0.130) Hematopoietic/Immune (0.130)	Cell Proliferation (0.650) Inflammation and Immune Response (0.310)	pINCY
38	Nervous (0.273) Gastrointestinal (0.182) Cardiovascular (0.136) Hematopoietic/Immune (0.136)	Cell Proliferation (0.682) Inflammation and Immune Response (0.273)	pINCY
39	Reproductive (0.296) Nervous (0.222) Gastrointestinal (0.130) Hematopoietic/Immune (0.130)	Cell Proliferation (0.593) Inflammation and Immune Response (0.352)	pINCY
40	Hematopoietic/Immune (0.286) Reproductive (0.286) Dermatologic (0.143) Gastrointestinal (0.143) Musculoskeletal (0.143)	Cell Proliferation (0.286) Inflammation and Immune Response (0.571)	pINCY
41	Nervous (0.231) Reproductive (0.205) Gastrointestinal (0.179) Hematopoietic/Immune (0.154) Cardiovascular (0.103)	Cell Proliferation (0.692) Inflammation and Immune Response (0.308)	pINCY
42	Nervous (0.263) Reproductive (0.228) Hematopoietic/Immune (0.158) Gastrointestinal (0.140)	Cell Proliferation (0.772) Inflammation and Immune Response (0.263)	PSPORT1

Table 3 (Cont.)

Seq ID NO:	Tissue Expression (Fraction of Total)	Disease Class (Fraction of Total)	Vector
43	Reproductive (0.269) Nervous (0.135) Gastrointestinal (0.129) Cardiovascular (0.105) Hematopoietic/Immune (0.105)	Cell Proliferation (0.702) Inflammation and Immune Response (0.327)	pINCY
44	Reproductive (0.247) Nervous (0.148) Gastrointestinal (0.136) Hematopoietic/Immune (0.136) Cardiovascular (0.123)	Cell Proliferation (0.593) Inflammation and Immune Response (0.346)	pINCY
45	Reproductive (0.255) Nervous (0.213) Gastrointestinal (0.149) Hematopoietic/Immune (0.106) Musculoskeletal (0.106)	Cell Proliferation (0.511) Inflammation and Immune Response (0.362)	pINCY
46	Reproductive (0.257) Cardiovascular (0.143) Gastrointestinal (0.143) Nervous (0.114)	Cell Proliferation (0.714) Inflammation and Immune Response (0.300)	pINCY
47	Reproductive (0.254) Gastrointestinal (0.159) Nervous (0.131) Hematopoietic/Immune (0.112) Cardiovascular (0.106)	Cell Proliferation (0.655) Inflammation and Immune Response (0.321)	pINCY
48	Reproductive (0.265) Hematopoietic/Immune (0.171) Gastrointestinal (0.145) Nervous (0.137)	Cell Proliferation (0.641) Inflammation and Immune Response (0.393)	pINCY
49	Reproductive (0.378) Gastrointestinal (0.178) Nervous (0.156)	Cell Proliferation (0.556) Inflammation and Immune Response (0.400)	pINCY
50	Reproductive (0.238) Gastrointestinal (0.160) Hematopoietic/Immune (0.132) Nervous (0.117) Cardiovascular (0.111)	Cell Proliferation (0.647) Inflammation and Immune Response (0.349)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
26	PITUNOT02	Library was constructed using RNA obtained from Clontech. The RNA was isolated from pituitary glands removed from a pool of 87 male and female donors, 15 to 75 years old.
27	SINTNOT13	Library was constructed using RNA isolated from ileum tissue obtained from a 25-year-old Asian female during a partial colectomy and temporary ileostomy. Pathology indicated moderately active chronic ulcerative colitis involving colonic mucosa from the distal margin to the ascending colon. Family history included hyperlipidemia, depressive disorder, malignant cervical neoplasm, and viral hepatitis A.
28	UTRSNOT02	Library was constructed using RNA isolated from uterine tissue removed from a 34-year-old Caucasian female during a vaginal hysterectomy. Patient history included mitral valve disorder. Family history included stomach cancer, congenital heart anomaly, irritable bowel syndrome, ulcerative colitis, colon cancer, cerebrovascular disease, type II diabetes, and depression.
29	OVARNOT09	Library was constructed using RNA isolated from ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Pathology indicated multiple follicular cysts ranging in size from 0.4 to 1.5 cm in the right and left ovaries, chronic cervicitis, and squamous metaplasia of the cervix. The endometrium was in weakly proliferative phase. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
30	ADRETUT06	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 57-year-old Caucasian female during a unilateral right adrenalectomy. Pathology indicated pheochromocytoma forming a nodular mass completely replacing the medulla of the adrenal gland.

Table 4 (Cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
31	HMC1NOT01	The library was constructed using RNA isolated from the HMC-1 human mast cell line derived from a 52-year-old female. Patient history included mast cell leukemia.
32	TBLYNOT01	The library was constructed at Stratagene (STR937214), using RNA isolated from a hybrid of T-B lymphoblasts from a leukemic cell line.
33	LUNGAST01	The library was constructed using RNA isolated from the lung tissue of a 17-year-old Caucasian male, who died from head trauma. Patient history included asthma.
34	CERVNOT01	The library was constructed using RNA isolated from the uterine cervical tissue of a 35-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated mild chronic cervicitis. Family history included atherosclerotic coronary artery disease and type II diabetes.
35	BLADTUT02	The library was constructed using RNA isolated from bladder tumor tissue removed from an 80-year-old Caucasian female during a radical cystectomy and lymph node excision. Pathology indicated grade 3 invasive transitional cell carcinoma. Family history included acute renal failure, osteoarthritis, and atherosclerosis.
36	THYRNOT03	The library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.

Table 4 (Cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
37	BLADTUT04	The library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Patient history included tobacco use. Family history included type I diabetes, a malignant neoplasm of the stomach, atherosclerotic coronary artery disease, and an acute myocardial infarction.
38	STOMTUT02	The library was constructed using RNA isolated from stomach tumor tissue obtained from a 68-year-old Caucasian female during a partial gastrectomy. Pathology indicated a malignant lymphoma of diffuse large-cell type. Previous surgeries included cholecystectomy. Patient history included thalassemia. Family history included acute leukemia, malignant neoplasm of the esophagus, malignant stomach neoplasm, and atherosclerotic coronary artery disease.
39	STOMTUT02	The library was constructed using RNA isolated from stomach tumor tissue obtained from a 68-year-old Caucasian female during a partial gastrectomy. Pathology indicated a malignant lymphoma of diffuse large-cell type. Previous surgeries included cholecystectomy. Patient history included thalassemia. Family history included acute leukemia, malignant neoplasm of the esophagus, malignant stomach neoplasm, and atherosclerotic coronary artery disease.
40	LEUKNOT03	The library was constructed using RNA isolated from white blood cells of a 27-year-old female with blood type A+. The donor tested negative for cytomegalovirus (CMV).
41	ISLTNOT01	The library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.

Table 4 (Cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
42	KIDNNOT05	The library was constructed using RNA isolated from the kidney tissue of a 2-day-old Hispanic female, who died from cerebral anoxia. Family history included congenital heart disease.
43	THP1NOT03	The library was constructed using 1 microgram of polyA RNA isolated from untreated THP-1 cells. THP-1 (ATCC TIB 202) is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).
44	BRSTNOT12	The library was constructed using RNA isolated from diseased breast tissue removed from a 32-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included benign hypertension and atherosclerotic coronary artery disease.
45	THYRNOT10	The library was constructed using RNA isolated from the diseased left thyroid tissue removed from a 30-year-old Caucasian female during a unilateral thyroid lobectomy and parathyroid reimplantation. Pathology indicated lymphocytic thyroiditis.
46	293TF5T01	The library was constructed using RNA isolated from a transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue transfected with bgal. The cells were transformed with adenovirus 5 DNA.
47	PENCNOT07	The library was constructed using RNA isolated from penis right corpora cavernosa tissue removed from a male.

Table 4 (Cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
48	BRSTNOT27	The library was constructed using RNA isolated from right breast tissue removed from a 57-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated benign fat replaced breast parenchyma. Pathology for the associated tumor adenocarcinoma and extensive grade 2 intraductal carcinoma. Multiple (9 of 19) axillary lymph nodes were positive for metastatic adenocarcinoma with minimal extranodal extension. The largest nodal metastasis measured less than 1 cm in greatest dimension. Immunoperoxidase stains for estrogen and progesterone receptors were positive. Patient history included benign hypertension, hyperlipidemia, cardiac dysrhythmia, a benign colon neoplasm, a solitary breast cyst, and a breast neoplasm of uncertain behavior. Previous surgeries included appendectomy. Family history included benign hypertension, acute leukemia, primary liver cancer, and upper lobe lung cancer.
49	BRSTNOT32	The library was constructed using RNA isolated from diseased right breast tissue removed from a 46-year-old Caucasian female during bilateral subcutaneous mammectomy. Pathology indicated nonproliferative fibrocystic disease. Family history included breast cancer, benign hypertension, and atherosclerotic coronary artery disease.
50	TLYMNOT08	The library was constructed using RNA isolated from anergic allogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male. The cells were incubated for 3 days in the presence of OKT3 mAb (tissue culture flasks coated with 1 microgram/ml OKT3) and 5% human serum. The patient had no allergies.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S F. et al (1990) J. Mol. Biol. 215:403-410; Altschul, S F. et al. (1997) Nucleic Acids Res. 25. 3389-3402.	*ESTs* Probability value= 1.0E-8 or less *Full Length sequences* Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J Lipman (1988) Proc. Natl. Acad Sci. 85.2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2.482-489.	*ESTs* fasta E value=1 06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater, fastx E value=1.0E-8 or less *Full Length sequences* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19 6565-72, 1991 J G Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T K. et al (1997) J. Chem. Inf Comput. Sci 37 417-424.	Score=1000 or greater, Ratio of Score/Strength = 0 75 or larger, and Probability value= 1 0E-3 or less where applicable
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al (1994) J Mol Biol , 235 1501-1531, Sonnhammer, E L L. et al (1988) Nucleic Acids Res 26 320-322.	Score=10-50 bits, depending on individual protein families

Table 5 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A et al. (1997) Nucleic Acids Res 25 217-221.	Score= 4 0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8.186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8.195-202	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J M and S. Audic (1997) CABIOS 12 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra, Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50 and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.

12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

18. A purified antagonist of the polypeptide of claim 1.

19. A method for treating or preventing a disorder associated with decreased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a disorder associated with increased expression or activity of RNAAP, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

M·H

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) **International Patent Classification** [7] :
C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, A61K 38/17, C12Q 1/68

A2

(11) **International Publication Number:** **WO 00/11171**

(43) **International Publication Date:** 2 March 2000 (02.03.00)

(21) **International Application Number:** PCT/US99/19361

(22) **International Filing Date:** 20 August 1999 (20.08.99)

(30) **Priority Data:**

60/097,550	21 August 1998 (21.08.98)	US
60/115,639	12 January 1999 (12.01.99)	US

(63) **Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications**

US	60/115,639 (CIP)
Filed on	12 January 1999 (12.01.99)
US	60/097,550 (CIP)
Filed on	21 August 1998 (21.08.98)

(71) **Applicant** *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). GORGONE, Gina, A. [US/US]; 1253 Pinecrest Drive, Boulder Creek, CA 95006 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Drive, Sunnyvale, CA 94086 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). SHIH, Leo, L. [US/US]; Apartment B, 1081 Tanland Drive, Palo Alto, CA 94304 (US). YANG, Junming [CN/US]; 7136 Clarendon Street, San Jose, CA 95129 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95126 (US).

(74) **Agents:** BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) **Designated States:** AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) **Title:** HUMAN RNA–ASSOCIATED PROTEINS

(57) **Abstract**

The invention provides human RNA–associated proteins (RNAAP) and polynucleotides which identify and encode RNAAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with the expression of RNAAP.

Docket No.: PF-0579 USN

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

HUMAN RNA PROCESSING PROTEINS

the specification of which:

/___/ is attached hereto.

/ X / was filed on February 14, 2001, as application Serial No. 09/763,233 and if this box contains an X /___/, was amended on _____.

/ X / was filed as Patent Cooperation Treaty international application No. PCT/US99/19361 on 20 August,1999, if this box contains an X /_/, was amended on under Patent Cooperation Treaty Article 19 on _____ 2001, and if this box contains an X /_/, was amended on _____ _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

Docket No.: PF-0579 USN

Country	Number	Filing Date	Priority Claimed
			/_/ Yes /_/ No
			/_/ Yes /_/ No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/097,550	21 August 1998	Expired
60/115, 639	12 January 1999	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

Lucy J. Billings	Reg. No. 36,749
Michael C. Cerrone	Reg. No. 39,132
Diana Hamlet-Cox	Reg. No. ~~33,302~~ 14
Richard C. Ekstrom	Reg. No. ~~37,027~~
Barrie D. Greene	Reg. No. ~~46,740~~
Matthew R. Kaser	Reg. No. ~~44,817~~
Lynn E. Murry	Reg. No. 42,918
Shirley A. Recipon	Reg. No. 47,016
Susan K. Sather	Reg. No. 44,316
Michelle M. Stempien	Reg. No. 41,327
David G. Streeter	Reg. No. 43,168
Stephen Todd	Reg. No. ~~47,139~~
Christopher Turner	Reg. No. ~~45,167~~
P. Ben Wang	Reg. No. 41,420

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0579 USN

LEGAL DEPARTMENT
INCYTE GENOMICS, INC.
3160 PORTER DRIVE, PALO ALTO, CA 94304

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Sole Inventor or First Joint Inventor: 1 of 2

Full name: JENNIFER L. HILLMAN

Signature:

Date: February 16 , 2001

Citizenship United States

Residence: Mountain View, California 94040 CA

P.O. Address: 230 Monroe Drive, #17
Mountain View, California 94040

Second Joint Inventor: 2 of 2

Full name: HENRY YUE

Signature:

Date: March 9 , 2001

Citizenship United States

Residence: 826 Lois Avenue
Sunnyvale, California 94087 CA

P.O. Address: Sunnyvale, California 94087

Docket No.: PF-0579 USN

Third Joint Inventor: 3-00

Full name:	Y. TOM TANG
Signature:	Y. Tom Tang
Date:	February 27, 2001
Citizenship	~~People's Republic of China~~ USA
Residence:	San Jose, California 95118 CA
P.O. Address:	4230 Ranwick Court San Jose, California 95118

Fourth Joint Inventor: 4-00

Full name:	NEIL C. CORLEY
Signature:	Neil C. Corley
Date:	MARCH 5, 2001
Citizenship	United States
Residence:	Castro Valley, California 94552 CA
P.O. Address:	20426 Crow Creek Road Castro Valley, California 94552

Fifth Joint Inventor: 5-00

Full name:	KARL J. GUEGLER
Signature:	[signature]
Date:	02/02, 2001
Citizenship	Switzerland
Residence:	Menlo Park, California 94025 CA
P.O. Address:	1048 Oakland Avenue Menlo Park, California 9 4025

Docket No.: PF-0579 USN

Sixth Joint Inventor: 6-OD

Full name:	**GINA A. GORGONE**
Signature:	
Date:	Feb 5, **2001**
Citizenship	**United States**
Residence:	**Boulder Creek, California 95006** CA
P.O. Address:	**1253 Pinecrest Drive** **Boulder Creek, California 95006**

Seventh Joint Inventor: 7-OD

Full name:	**CHANDRA PATTERSON**
Signature:	
Date:	February 7, **2001**
Citizenship	**United States**
Residence:	**Menlo Park, California 94025** CA
P.O. Address:	**490 Sherwood Way, #1** **Menlo Park, California 94025**

Eighth Joint Inventor: 8-OD

Full name:	**MARIAH R. BAUGHN**
Signature:	
Date:	February 12, **2001**
Citizenship	**United States**
Residence:	**Boulder Creek, California 94577** CA
P.O. Address:	**14244 Santiago Road** **San Leandro, California 94577**

Ninth Joint Inventor:

9-00

Full name:	**PREETI LAL**
Signature:	Preeti Lal
Date:	FEBRUARY, 16 , 2001
Citizenship	**India**
Residence:	**Santa Clara, California 95056** CA
P.O. Address:	**P.O. Box 5142** **Santa Clara, California 95056**

Tenth Joint Inventor:

10-00

Full name:	**OLGA BANDMAN**
Signature:	Olga Bandman
Date:	16 February , 2001
Citizenship	**United States**
Residence:	**Mountain View, California 94043** CA
P.O. Address:	**366 Anna Avenue** **Mountain View, California 94043**

Eleventh Joint Inventor:

11-00

Full name:	**ROOPA REDDY**
Signature:	Roopa Reddy
Date:	March 1st , 2001
Citizenship	**India**
Residence:	**Sunnyvale, California 94086** CA
P.O. Address:	**1233 W. McKinley Avenue, #3** **Sunnyvale, California 94086**

Docket No.: PF-0579 USN

Twelfth Joint Inventor:

12-00

Full name:	YALDA AZIMZAI
Signature:	Yalda Azimzai
Date:	February 2nd, 2001
Citizenship	United States
Residence:	Castro Valley, California 94552 CA
P.O. Address:	5518 Boulder Canyon Drive Castro Valley, California 94552

Thirteenth Joint Inventor:

13-00

Full name:	LEO L. SHIH
Signature:	[signature]
Date:	March 12, 2001
Citizenship	United States
Residence:	~~Palo Alto~~ East Palo Alto, California 94303 CA
P.O. Address:	~~1081 Tanland Drive, Apt. B~~ ~~Palo Alto, California 94303~~ 908 O'Connor St. East Palo Alto, CA 94303

L.S.
3/12/2001

Fourteenth Joint Inventor:

14-00

Full name:	JUNMING YANG
Signature:	[signature]
Date:	February 16, 2001
Citizenship	China
Residence:	San Jose, California 95129 CA
P.O. Address:	7125 Bark Lane San Jose, California 95129

Docket No.: PF-0579 USN

Fifteenth Joint Inventor:

15th

Full name:	DYUNG AINA M. LU
Signature:	
Date:	March 22, 2001
Citizenship	United States
Residence:	San Jose, California 95123 CA
P.O. Address:	233 Coy Drive San Jose, California 95123

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
HILLMAN, Jennifer L.
YUE, Henry
TANG, Y. Tom
CORLEY, Neil C.
GUEGLER, Karl J.
GORGONE, Gina A.
PATTERSON, Chandra
BAUGHN, Mariah R.
LAL, Preeti
BANDMAN, Olga
REDDY, Roopa
AZIMZAI, Yalda
SHIH, Leo L.
YANG, Junming
LU, Dyung Aina M.

<120> HUMAN RNA-ASSOCIATED PROTEINS

<130> PF-0579 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/097,550; 60/115,639
<151> 1998-08-21; 1999-01-12

<160> 50

<170> PERL Program

<210> 1
<211> 216
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 399781CD1

<400> 1

```
Met Ser Arg Tyr Leu Arg Pro Pro Asn Thr Ser Leu Phe Val Arg
  1               5                  10                  15
Asn Val Ala Asp Asp Thr Arg Ser Glu Asp Leu Arg Arg Glu Phe
                20                  25                  30
Gly Arg Tyr Gly Pro Ile Val Asp Val Tyr Val Pro Leu Asp Phe
                35                  40                  45
Tyr Thr Arg Arg Pro Arg Gly Phe Ala Tyr Val Gln Phe Glu Asp
                50                  55                  60
Val Arg Asp Ala Glu Asp Ala Leu His Asn Leu Asp Arg Lys Trp
                65                  70                  75

Ile Cys Gly Arg Gln Ile Glu Ile Gln Phe Ala Gln Gly Asp Arg
                80                  85                  90
```

```
Lys Thr Pro Asn Gln Met Lys Ala Lys Glu Gly Arg Asn Val Tyr
                 95                  100                 105
Ser Ser Ser Arg Tyr Asp Asp Tyr Asp Arg Tyr Arg Arg Ser Arg
                110                 115                 120
Ser Arg Ser Tyr Glu Arg Arg Arg Ser Arg Ser Arg Ser Phe Asp
                125                 130                 135
Tyr Asn Tyr Arg Arg Ser Tyr Ser Pro Arg Asn Ser Arg Pro Thr
                140                 145                 150
Gly Arg Pro Arg Arg Arg Glu Ala Ile Pro Thr Met Ile Asp Gln
                155                 160                 165
Thr Ala Ala Gly Ile Pro Ser Thr Val Leu Leu Thr Thr Leu Gln
                170                 175                 180
Glu Arg Ser Glu Ser Gly Lys Arg Thr Lys Glu Gly Gln Phe Lys
                185                 190                 195
Arg Pro Lys Gly Gly Trp Lys Val Leu Gln Tyr Glu Tyr Cys Thr
                200                 205                 210
Asn Ile Leu Thr Leu Val
                215
```

<210> 2
<211> 962
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1806542CD1

<400> 2

```
Met Asp Glu Gln Ala Leu Leu Gly Leu Asn Pro Asn Ala Asp Ser
  1               5                  10                  15
Asp Phe Arg Gln Arg Ala Leu Ala Tyr Phe Glu Gln Leu Lys Ile
                20                  25                  30
Ser Pro Asp Ala Trp Gln Val Cys Ala Glu Ala Leu Ala Gln Arg
                35                  40                  45
Thr Tyr Ser Asp Asp His Val Lys Phe Phe Cys Phe Gln Val Leu
                50                  55                  60
Glu His Gln Val Lys Tyr Lys Tyr Ser Glu Leu Thr Thr Val Gln
                65                  70                  75
Gln Gln Leu Ile Arg Glu Thr Leu Ile Ser Trp Leu Gln Ala Gln
                80                  85                  90
Met Leu Asn Pro Gln Pro Glu Lys Thr Phe Ile Arg Asn Lys Ala
                95                 100                 105
Ala Gln Val Phe Ala Leu Leu Phe Val Thr Glu Tyr Leu Thr Lys
               110                 115                 120
Trp Pro Lys Phe Phe Phe Asp Ile Leu Ser Val Val Asp Leu Asn
               125                 130                 135
Pro Arg Gly Val Asp Leu Tyr Leu Arg Ile Leu Met Ala Ile Asp
               140                 145                 150
Ser Glu Leu Val Asp Arg Asp Val Val His Thr Ser Glu Glu Ala
               155                 160                 165
Arg Arg Asn Thr Leu Ile Lys Asp Thr Met Arg Glu Gln Cys Ile
               170                 175                 180
Pro Asn Leu Val Glu Ser Trp Tyr Gln Ile Leu Gln Asn Tyr Gln
               185                 190                 195
Phe Thr Asn Ser Glu Val Thr Cys Gln Cys Leu Glu Val Val Gly
```

```
                200                 205                 210
Ala Tyr Val Ser Trp Ile Asp Leu Ser Leu Ile Ala Asn Asp Arg
                215                 220                 225
Phe Ile Asn Met Leu Leu Gly His Met Ser Ile Glu Val Leu Arg
                230                 235                 240
Glu Glu Ala Cys Asp Cys Leu Phe Glu Val Val Asn Lys Gly Met
                245                 250                 255
Asp Pro Val Asp Lys Met Lys Leu Val Glu Ser Leu Cys Gln Val
                260                 265                 270
Leu Gln Ser Ala Gly Phe Phe Ser Ile Asp Gln Glu Glu Asp Val
                275                 280                 285
Asp Phe Leu Ala Arg Phe Ser Lys Leu Val Asn Gly Met Gly Gln
                290                 295                 300
Ser Leu Ile Val Ser Trp Ser Lys Leu Ile Lys Asn Gly Asp Ile
                305                 310                 315
Lys Asn Ala Gln Glu Ala Leu Gln Ala Ile Glu Thr Lys Val Ala
                320                 325                 330
Leu Met Leu Gln Leu Leu Ile His Glu Asp Asp Ile Ser Ser
                335                 340                 345
Asn Ile Ile Gly Phe Cys Tyr Asp Tyr Leu His Ile Leu Lys Gln
                350                 355                 360
Leu Thr Val Leu Ser Asp Gln Gln Lys Ala Asn Val Glu Ala Ile
                365                 370                 375
Met Leu Ala Val Met Lys Lys Leu Thr Tyr Asp Glu Glu Tyr Asn
                380                 385                 390
Phe Glu Asn Glu Gly Glu Asp Glu Ala Met Phe Val Glu Tyr Arg
                395                 400                 405
Lys Gln Leu Lys Leu Leu Leu Asp Arg Leu Ala Gln Val Ser Pro
                410                 415                 420
Glu Leu Leu Leu Ala Ser Val Arg Arg Val Phe Ser Ser Thr Leu
                425                 430                 435
Gln Asn Trp Gln Thr Thr Arg Phe Met Glu Val Glu Val Ala Ile
                440                 445                 450
Arg Leu Leu Tyr Met Leu Ala Glu Ala Leu Pro Val Ser His Gly
                455                 460                 465
Ala His Phe Ser Gly Asp Val Ser Lys Ala Ser Ala Leu Gln Asp
                470                 475                 480
Met Met Arg Thr Leu Val Thr Ser Gly Val Ser Ser Tyr Gln His
                485                 490                 495
Thr Ser Val Thr Leu Glu Phe Phe Glu Thr Val Val Arg Tyr Glu
                500                 505                 510
Lys Phe Phe Thr Val Glu Pro Gln His Ile Pro Cys Val Leu Met
                515                 520                 525
Ala Phe Leu Asp His Arg Gly Leu Arg His Ser Ser Ala Lys Val
                530                 535                 540
Arg Ser Arg Thr Ala Tyr Leu Phe Ser Arg Phe Val Lys Ser Leu
                545                 550                 555
Asn Lys Gln Met Asn Pro Phe Ile Glu Asp Ile Leu Asn Arg Ile
                560                 565                 570
Gln Asp Leu Leu Glu Leu Ser Pro Pro Glu Asn Gly His Gln Ser
                575                 580                 585
Leu Leu Ser Ser Asp Asp Gln Leu Phe Ile Tyr Glu Thr Ala Gly
                590                 595                 600
Val Leu Ile Val Asn Ser Glu Tyr Pro Ala Glu Arg Lys Gln Ala
                605                 610                 615
Leu Met Arg Asn Leu Leu Thr Pro Leu Met Glu Lys Phe Lys Ile
                620                 625                 630
```

```
Leu Leu Glu Lys Leu Met Leu Ala Gln Asp Glu Glu Arg Gln Ala
                635                 640                 645
Ser Leu Ala Asp Cys Leu Asn His Ala Val Gly Phe Ala Ser Arg
                650                 655                 660
Thr Ser Lys Ala Phe Ser Asn Lys Gln Thr Val Lys Gln Cys Gly
                665                 670                 675
Cys Ser Glu Val Tyr Leu Asp Cys Leu Gln Thr Phe Leu Pro Ala
                680                 685                 690
Leu Ser Cys Pro Leu Gln Lys Asp Ile Leu Arg Ser Gly Val Arg
                695                 700                 705
Thr Phe Leu His Arg Met Ile Ile Cys Leu Glu Glu Glu Val Leu
                710                 715                 720
Pro Phe Ile Pro Ser Ala Ser Glu His Met Leu Lys Asp Cys Glu
                725                 730                 735
Ala Lys Asp Leu Gln Glu Phe Ile Pro Leu Ile Asn Gln Ile Thr
                740                 745                 750
Ala Lys Phe Lys Ile Gln Val Ser Pro Phe Leu Gln Gln Met Phe
                755                 760                 765
Met Pro Leu Leu His Ala Ile Phe Glu Val Leu Leu Arg Pro Ala
                770                 775                 780
Glu Glu Asn Asp Gln Ser Ala Ala Leu Glu Lys Gln Met Leu Arg
                785                 790                 795
Arg Ser Tyr Phe Ala Phe Leu Gln Thr Val Thr Gly Ser Gly Met
                800                 805                 810
Ser Glu Val Ile Ala Asn Gln Gly Ala Glu Asn Val Glu Arg Val
                815                 820                 825
Leu Val Thr Val Ile Gln Gly Ala Val Glu Tyr Pro Asp Pro Ile
                830                 835                 840
Ala Gln Lys Thr Cys Phe Ile Ile Leu Ser Lys Leu Val Glu Leu
                845                 850                 855
Trp Gly Gly Lys Asp Gly Pro Val Gly Phe Ala Asp Phe Val Tyr
                860                 865                 870
Lys His Ile Val Pro Ala Cys Phe Leu Ala Pro Leu Lys Gln Thr
                875                 880                 885
Phe Asp Leu Ala Asp Ala Gln Thr Val Leu Ala Leu Ser Glu Cys
                890                 895                 900
Ala Val Thr Leu Lys Thr Ile His Leu Lys Arg Gly Pro Glu Cys
                905                 910                 915
Val Gln Tyr Leu Gln Gln Glu Tyr Leu Pro Ser Leu Gln Val Ala
                920                 925                 930
Pro Glu Ile Ile Gln Glu Phe Cys Gln Ala Leu Gln Gln Pro Asp
                935                 940                 945
Ala Lys Val Phe Lys Asn Tyr Leu Lys Val Phe Phe Gln Arg Ala
                950                 955                 960
Lys Pro
```

<210> 3
<211> 285
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2263514CD1

<400> 3
Met Asp Trp Val Met Lys His Asn Gly Pro Asn Asp Ala Ser Asp
1 5 10 15
Gly Thr Val Arg Leu Arg Gly Leu Pro Phe Gly Cys Ser Lys Glu
20 25 30
Glu Ile Val Arg Val Leu Ser Arg Tyr Ile Glu Ile Phe Arg Ser
35 40 45
Ser Arg Ser Glu Ile Lys Gly Phe Tyr Asp Pro Pro Arg Arg Leu
50 55 60
Leu Gly Gln Arg Pro Gly Pro Tyr Asp Arg Pro Ile Gly Gly Arg
65 70 75
Gly Gly Tyr Tyr Gly Ala Gly Arg Gly Ser Tyr Gly Gly Phe Asp
80 85 90
Asp Tyr Gly Gly Tyr Asn Asn Tyr Gly Tyr Gly Asn Asp Gly Phe
95 100 105
Asp Asp Arg Met Arg Asp Gly Arg Gly Met Gly Gly His Gly Tyr
110 115 120
Gly Gly Ala Gly Asp Ala Ser Ser Gly Phe His Gly Gly His Phe
125 130 135
Val His Met Arg Gly Leu Pro Phe Arg Ala Thr Glu Asn Ala Ile
140 145 150
Ala Asn Phe Phe Ser Pro Leu Asn Pro Ile Arg Val His Ile Asp
155 160 165
Ile Gly Ala Asp Gly Arg Ala Thr Gly Glu Ala Asp Val Glu Phe
170 175 180
Val Thr His Glu Asp Ala Val Ala Ala Met Ser Lys Asp Lys Asn
185 190 195
Asn Met Gln His Arg Tyr Ile Glu Leu Phe Leu Asn Ser Thr Pro
200 205 210
Gly Gly Gly Ser Gly Met Gly Gly Ser Gly Met Gly Gly Tyr Gly
215 220 225
Arg Asp Gly Met Asp Asn Gln Gly Gly Tyr Gly Ser Val Gly Arg
230 235 240
Met Gly Met Gly Asn Asn Tyr Ser Gly Gly Tyr Gly Thr Pro Asp
245 250 255
Gly Leu Gly Gly Tyr Gly Arg Gly Gly Gly Gly Ser Gly Gly Tyr
260 265 270
Tyr Gly Gln Gly Gly Met Ser Gly Gly Gly Trp Arg Gly Met Tyr
275 280 285

<210> 4
<211> 267
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2738270CD1

<400> 4
Met Gly Ala Ala Ala Ala Glu Ala Asp Arg Thr Leu Phe Val Gly
1 5 10 15
Asn Leu Glu Thr Lys Val Thr Glu Glu Leu Leu Phe Glu Leu Phe
20 25 30
His Gln Ala Gly Pro Val Ile Lys Val Lys Ile Pro Lys Asp Lys
35 40 45

```
Asp Gly Lys Pro Lys Gln Phe Ala Phe Val Asn Phe Lys His Glu
                  50                  55                  60
Val Ser Val Pro Tyr Ala Met Asn Leu Leu Asn Gly Ile Lys Leu
                  65                  70                  75
Tyr Gly Arg Pro Ile Lys Ile Gln Phe Arg Ser Gly Ser Ser His
                  80                  85                  90
Ala Pro Gln Asp Val Ser Leu Ser Tyr Pro Gln His His Val Gly
                  95                 100                 105
Asn Ser Ser Pro Thr Ser Thr Ser Pro Ser Ser Arg Tyr Glu Arg
                 110                 115                 120
Thr Met Asp Asn Met Thr Ser Ser Ala Gln Ile Ile Gln Arg Ser
                 125                 130                 135
Phe Ser Ser Pro Glu Asn Phe Gln Arg Gln Ala Val Met Asn Ser
                 140                 145                 150
Ala Leu Arg Gln Met Ser Tyr Gly Gly Lys Phe Gly Ser Ser Pro
                 155                 160                 165
Leu Asp Gln Ser Gly Phe Ser Pro Ser Val Gln Ser His Ser His
                 170                 175                 180
Ser Phe Asn Gln Ser Ser Ser Ser Gln Trp Arg Gln Gly Thr Pro
                 185                 190                 195
Ser Ser Gln Arg Lys Val Arg Met Asn Ser Tyr Pro Tyr Leu Ala
                 200                 205                 210
Asp Arg His Tyr Ser Arg Glu Gln Arg Tyr Thr Asp His Gly Ser
                 215                 220                 225
Asp His His Tyr Arg Gly Lys Arg Asp Asp Phe Phe Tyr Glu Asp
                 230                 235                 240
Arg Asn His Asp Asp Trp Ser His Asp Tyr Asp Asn Arg Arg Asp
                 245                 250                 255
Ser Ser Arg Asp Gly Lys Trp Arg Ser Ser Arg His
                 260                 265
```

<210> 5
<211> 369
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2824412CD1

<400> 5

```
Met Pro Pro Gln Pro Gln Gly Pro Ala Pro Leu Arg Arg Pro Asp
  1               5                  10                  15
Ser Ser Asp Asp Arg Tyr Val Met Thr Lys His Ala Thr Ile Tyr
                 20                  25                  30
Pro Thr Glu Glu Glu Leu Gln Ala Val Gln Lys Ile Val Ser Ile
                 35                  40                  45
Thr Glu Arg Ala Leu Lys Leu Val Ser Asp Ser Leu Ser Glu His
                 50                  55                  60
Glu Lys Asn Lys Asn Lys Glu Gly Asp Asp Lys Lys Glu Gly Gly
                 65                  70                  75
Lys Asp Arg Ala Leu Lys Gly Val Leu Arg Val Gly Val Leu Ala
                 80                  85                  90
Lys Gly Leu Leu Leu Arg Gly Asp Arg Asn Val Asn Leu Val Leu
```

```
                 95                  100                 105
Leu Cys Ser Glu Lys Pro Ser Lys Thr Leu Leu Ser Arg Ile Ala
                110                 115                 120
Glu Asn Leu Pro Lys Gln Leu Ala Val Ile Ser Pro Glu Lys Tyr
                125                 130                 135
Asp Ile Lys Cys Ala Val Ser Glu Ala Ala Ile Ile Leu Asn Ser
                140                 145                 150
Cys Val Glu Pro Lys Met Gln Val Thr Ile Thr Leu Thr Ser Pro
                155                 160                 165
Ile Ile Arg Glu Glu Asn Met Arg Glu Gly Asp Val Thr Ser Gly
                170                 175                 180
Met Val Lys Asp Pro Pro Asp Val Leu Asp Arg Gln Lys Cys Leu
                185                 190                 195
Asp Ala Leu Ala Ala Leu Arg His Ala Lys Trp Phe Gln Ala Arg
                200                 205                 210
Ala Asn Gly Leu Gln Ser Cys Val Ile Ile Ile Arg Ile Leu Arg
                215                 220                 225
Asp Leu Cys Gln Arg Val Pro Thr Trp Ser Asp Phe Pro Ser Trp
                230                 235                 240
Ala Met Glu Leu Leu Val Glu Lys Ala Ile Ser Ser Ala Ser Ser
                245                 250                 255
Pro Gln Ser Pro Gly Asp Ala Leu Arg Arg Val Phe Glu Cys Ile
                260                 265                 270
Ser Ser Gly Ile Ile Leu Lys Gly Ser Pro Gly Leu Leu Asp Pro
                275                 280                 285
Cys Glu Lys Asp Pro Phe Asp Thr Leu Ala Thr Met Thr Asp Gln
                290                 295                 300
Gln Arg Glu Asp Ile Thr Ser Ser Ala Gln Phe Ala Leu Arg Leu
                305                 310                 315
Leu Ala Phe Arg Gln Ile His Lys Val Leu Gly Met Asp Pro Leu
                320                 325                 330
Pro Gln Met Ser Gln Arg Phe Asn Ile His Asn Asn Arg Lys Arg
                335                 340                 345
Arg Arg Asp Ser Asp Gly Val Asp Gly Phe Glu Ala Glu Gly Lys
                350                 355                 360
Lys Asp Lys Lys Asp Tyr Asp Asn Phe
                365
```

<210> 6
<211> 175
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 002690CD1

<400> 6

```
Met Arg Leu Ser Val Ala Ala Ala Ile Ser His Gly Arg Val Phe
  1               5                  10                  15
Arg Arg Met Gly Leu Gly Pro Glu Ser Arg Ile His Leu Leu Arg
                20                  25                  30
Asn Leu Leu Thr Gly Leu Val Arg His Glu Arg Ile Glu Ala Pro
                35                  40                  45
```

```
Trp Ala Arg Val Asp Glu Met Arg Gly Tyr Ala Glu Lys Leu Ile
                 50                  55                  60
Asp Tyr Gly Lys Leu Gly Asp Thr Asn Glu Arg Ala Met Arg Met
                 65                  70                  75
Ala Asp Phe Trp Leu Thr Glu Lys Asp Leu Ile Pro Lys Leu Phe
                 80                  85                  90
Gln Val Leu Ala Pro Arg Tyr Lys Asp Gln Thr Gly Gly Tyr Thr
                 95                 100                 105
Arg Met Leu Gln Ile Pro Asn Arg Ser Leu Asp Arg Ala Lys Met
                110                 115                 120
Ala Val Ile Glu Tyr Lys Gly Asn Cys Leu Pro Pro Leu Pro Leu
                125                 130                 135
Pro Arg Arg Asp Ser His Leu Thr Leu Leu Asn Gln Leu Leu Gln
                140                 145                 150
Gly Leu Arg Gln Asp Leu Arg Gln Ser Gln Glu Ala Ser Asn His
                155                 160                 165
Ser Ser His Thr Ala Gln Thr Pro Gly Ile
                170                 175
```

<210> 7
<211> 311
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 041108CD1

<400> 7

```
Met Leu Gln Phe Val Arg Ala Gly Ala Arg Ala Trp Leu Arg Pro
  1                  5                  10                  15
Thr Gly Ser Gln Gly Leu Ser Ser Leu Ala Glu Glu Ala Ala Arg
                 20                  25                  30
Ala Thr Glu Asn Pro Glu Gln Val Ala Ser Glu Gly Leu Pro Glu
                 35                  40                  45
Pro Val Leu Arg Lys Val Glu Leu Pro Val Pro Thr His Arg Arg
                 50                  55                  60
Pro Val Gln Ala Trp Val Glu Ser Leu Arg Gly Phe Glu Gln Glu
                 65                  70                  75
Arg Val Gly Leu Ala Asp Leu His Pro Asp Val Phe Ala Thr Ala
                 80                  85                  90
Pro Arg Leu Asp Ile Leu His Gln Val Ala Met Trp Gln Lys Asn
                 95                 100                 105
Phe Lys Arg Ile Ser Tyr Ala Lys Thr Lys Thr Arg Ala Glu Val
                110                 115                 120
Arg Gly Gly Gly Arg Lys Pro Trp Pro Gln Lys Gly Thr Gly Arg
                125                 130                 135
Ala Arg His Gly Ser Ile Arg Ser Pro Leu Trp Arg Gly Gly Gly
                140                 145                 150
Val Ala His Gly Pro Arg Gly Pro Thr Ser Tyr Tyr Tyr Met Leu
                155                 160                 165
Pro Met Lys Val Arg Ala Leu Gly Leu Lys Val Ala Leu Thr Val
                170                 175                 180
Lys Leu Ala Gln Asp Asp Leu His Ile Met Asp Ser Leu Glu Leu
```

```
                185                 190                 195
Pro Thr Gly Asp Pro Gln Tyr Leu Thr Glu Leu Ala His Tyr Arg
                200                 205                 210
Arg Trp Gly Asp Ser Val Leu Leu Val Asp Leu Thr His Glu Glu
                215                 220                 225
Met Pro Gln Ser Ile Val Glu Ala Thr Ser Arg Leu Lys Thr Phe
                230                 235                 240
Asn Leu Ile Pro Ala Val Gly Leu Asn Val His Ser Met Leu Lys
                245                 250                 255
His Gln Thr Leu Val Leu Thr Leu Pro Thr Val Ala Phe Leu Glu
                260                 265                 270
Asp Lys Leu Leu Trp Gln Asp Ser Arg Tyr Arg Pro Leu Tyr Pro
                275                 280                 285
Phe Ser Leu Pro Tyr Ser Asp Phe Pro Arg Pro Leu Pro His Ala
                290                 295                 300
Thr Gln Gly Pro Ala Ala Thr Pro Tyr His Cys
                305                 310
```

<210> 8
<211> 330
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 869138CD1

<400> 8

```
Met Ser Thr Lys Asn Phe Arg Val Ser Asp Gly Asp Trp Ile Cys
  1                 5                  10                  15
Pro Asp Lys Lys Cys Gly Asn Val Asn Phe Ala Arg Arg Thr Ser
                 20                  25                  30
Cys Asn Arg Cys Gly Arg Glu Lys Thr Thr Glu Ala Lys Met Met
                 35                  40                  45
Lys Ala Gly Gly Thr Glu Ile Gly Lys Thr Leu Ala Glu Lys Ser
                 50                  55                  60
Arg Gly Leu Phe Ser Ala Asn Asp Trp Gln Cys Lys Thr Cys Ser
                 65                  70                  75
Asn Val Asn Trp Ala Arg Arg Ser Glu Cys Asn Met Cys Asn Thr
                 80                  85                  90
Pro Lys Tyr Ala Lys Leu Glu Glu Arg Thr Gly Tyr Gly Gly Gly
                 95                 100                 105
Phe Asn Glu Arg Glu Asn Val Glu Tyr Ile Glu Arg Glu Glu Ser
                110                 115                 120
Asp Gly Glu Tyr Asp Glu Phe Gly Arg Lys Lys Lys Lys Tyr Arg
                125                 130                 135
Gly Lys Ala Val Gly Pro Ala Ser Ile Leu Lys Glu Val Glu Asp
                140                 145                 150
Lys Glu Ser Glu Gly Glu Glu Glu Asp Glu Asp Glu Asp Leu Ser
                155                 160                 165
Lys Tyr Lys Leu Asp Glu Asp Glu Asp Glu Asp Asp Ala Asp Leu
                170                 175                 180
Ser Lys Tyr Asn Leu Asp Ala Ser Glu Glu Glu Asp Ser Asn Lys
                185                 190                 195
```

```
Lys Lys Ser Asn Arg Arg Ser Arg Ser Lys Ser Arg Ser Ser His
                200                 205                 210
Ser Arg Ser Ser Ser Arg Ser Ser Ser Pro Ser Ser Ser Arg Ser
                215                 220                 225
Arg Ser Arg Ser Arg Ser Arg Ser Ser Ser Ser Ser Gln Ser Arg
                230                 235                 240
Ser Arg Ser Ser Ser Arg Glu Arg Ser Arg Ser Arg Gly Ser Lys
                245                 250                 255
Ser Arg Ser Ser Ser Arg Ser His Arg Gly Ser Ser Ser Pro Arg
                260                 265                 270
Lys Arg Ser Tyr Ser Ser Ser Ser Ser Ser Pro Glu Arg Asn Arg
                275                 280                 285
Lys Arg Ser Arg Ser Arg Ser Ser Ser Ser Gly Asp Arg Lys Lys
                290                 295                 300
Arg Arg Thr Arg Ser Arg Ser Pro Glu Arg Arg His Arg Ser Ser
                305                 310                 315
Ser Gly Ser Ser His Ser Gly Ser Arg Ser Ser Ser Lys Lys Lys
                320                 325                 330
```

<210> 9
<211> 183
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 934406CD1

<400> 9

```
Met Ser Arg Tyr Leu Arg Pro Pro Asn Thr Ser Leu Phe Val Arg
  1               5                  10                  15
Asn Val Ala Asp Asp Thr Arg Ser Glu Asp Leu Arg Arg Glu Phe
                20                  25                  30
Gly Arg Tyr Gly Pro Ile Val Asp Val Tyr Val Pro Leu Asp Phe
                35                  40                  45
Tyr Thr Arg Arg Pro Arg Gly Phe Ala Tyr Val Gln Phe Glu Asp
                50                  55                  60
Val Arg Asp Ala Glu Asp Ala Leu His Asn Leu Asp Arg Lys Trp
                65                  70                  75
Ile Cys Gly Arg Gln Ile Glu Ile Gln Phe Ala Gln Gly Asp Arg
                80                  85                  90
Lys Thr Pro Asn Gln Met Lys Ala Lys Glu Gly Arg Asn Val Tyr
                95                 100                 105
Ser Ser Ser Arg Tyr Asp Asp Tyr Asp Arg Tyr Arg Arg Ser Arg
               110                 115                 120
Ser Arg Ser Tyr Glu Arg Arg Arg Ser Arg Ser Arg Ser Phe Asp
               125                 130                 135
Tyr Asn Tyr Arg Arg Ser Tyr Ser Pro Arg Asn Ser Arg Pro Thr
               140                 145                 150
Gly Arg Pro Arg Arg Ser Arg Ser His Ser Asp Asn Asp Arg Pro
               155                 160                 165
Asn Cys Ser Trp Asn Thr Gln Tyr Ser Ser Ala Tyr Tyr Thr Ser
               170                 175                 180
Arg Lys Ile
```

<210> 10
<211> 670
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1315083CD1

<400> 10

```
Met Ser His Leu Pro Met Lys Leu Leu Arg Lys Lys Ile Glu Lys
  1               5                  10                  15
Arg Asn Leu Lys Leu Arg Gln Arg Asn Leu Lys Phe Gln Gly Ala
             20                  25                  30
Ser Asn Leu Thr Leu Ser Glu Thr Gln Asn Gly Asp Val Ser Glu
             35                  40                  45
Glu Thr Met Gly Ser Arg Lys Val Lys Lys Ser Lys Gln Lys Pro
             50                  55                  60
Met Asn Val Gly Leu Ser Glu Thr Gln Asn Gly Gly Met Ser Gln
             65                  70                  75
Glu Ala Val Gly Asn Ile Lys Val Thr Lys Ser Pro Gln Lys Ser
             80                  85                  90
Thr Val Leu Thr Asn Gly Glu Ala Ala Met Gln Ser Ser Asn Ser
             95                 100                 105
Glu Ser Lys Lys Lys Lys Lys Lys Lys Arg Lys Met Val Asn Asp
            110                 115                 120
Ala Glu Pro Asp Thr Lys Lys Ala Lys Thr Glu Asn Lys Gly Lys
            125                 130                 135
Ser Glu Glu Glu Ser Ala Glu Thr Thr Lys Glu Thr Glu Asn Asn
            140                 145                 150
Val Glu Lys Pro Asp Asn Asp Glu Asp Glu Ser Glu Val Pro Ser
            155                 160                 165
Leu Pro Leu Gly Leu Thr Gly Ala Phe Glu Asp Thr Ser Phe Ala
            170                 175                 180
Ser Leu Cys Asn Leu Val Asn Glu Asn Thr Leu Lys Ala Ile Lys
            185                 190                 195
Glu Met Gly Phe Thr Asn Met Thr Glu Ile Gln His Lys Ser Ile
            200                 205                 210
Arg Pro Leu Leu Glu Gly Arg Asp Leu Leu Ala Ala Ala Lys Thr
            215                 220                 225
Gly Ser Gly Lys Thr Leu Ala Phe Leu Ile Pro Ala Val Glu Leu
            230                 235                 240
Ile Val Lys Leu Arg Phe Met Pro Arg Asn Gly Thr Gly Val Leu
            245                 250                 255
Ile Leu Ser Pro Thr Arg Glu Leu Ala Met Gln Thr Phe Gly Val
            260                 265                 270
Leu Lys Glu Leu Met Thr His His Val His Thr Tyr Gly Leu Ile
            275                 280                 285
Met Gly Gly Ser Asn Arg Ser Ala Glu Ala Gln Lys Leu Gly Asn
            290                 295                 300
Gly Ile Asn Ile Ile Val Ala Thr Pro Gly Arg Leu Leu Asp His
            305                 310                 315
Met Gln Asn Thr Pro Gly Phe Met Tyr Lys Asn Leu Gln Cys Leu
            320                 325                 330
```

```
Val Ile Asp Glu Ala Asp Arg Ile Leu Asp Val Gly Phe Glu Glu
                335                 340                 345
Glu Leu Lys Gln Ile Ile Lys Leu Leu Pro Thr Arg Arg Gln Thr
                350                 355                 360
Met Leu Phe Ser Ala Thr Gln Thr Arg Lys Val Glu Asp Leu Ala
                365                 370                 375
Arg Ile Ser Leu Lys Lys Glu Pro Leu Tyr Val Gly Val Asp Asp
                380                 385                 390
Asp Lys Ala Asn Ala Thr Val Asp Gly Leu Glu Gln Gly Tyr Val
                395                 400                 405
Val Cys Pro Ser Glu Lys Arg Phe Leu Leu Leu Phe Thr Phe Leu
                410                 415                 420
Lys Lys Asn Arg Lys Lys Lys Leu Met Val Phe Phe Ser Ser Cys
                425                 430                 435
Met Ser Val Lys Tyr His Tyr Glu Leu Leu Asn Tyr Ile Asp Leu
                440                 445                 450
Pro Val Leu Ala Ile His Gly Lys Gln Lys Gln Asn Lys Arg Thr
                455                 460                 465
Thr Thr Phe Phe Gln Phe Cys Asn Ala Asp Ser Gly Thr Leu Leu
                470                 475                 480
Cys Thr Asp Val Ala Ala Arg Gly Leu Asp Ile Pro Glu Val Asp
                485                 490                 495
Trp Ile Val Gln Tyr Asp Pro Pro Asp Asp Pro Lys Glu Tyr Ile
                500                 505                 510
His Arg Val Gly Arg Thr Ala Arg Gly Leu Asn Gly Arg Gly His
                515                 520                 525
Ala Leu Leu Ile Leu Arg Pro Glu Glu Leu Gly Phe Leu Arg Tyr
                530                 535                 540
Leu Lys Gln Ser Lys Val Pro Leu Ser Glu Phe Asp Phe Ser Trp
                545                 550                 555
Ser Lys Ile Ser Asp Ile Gln Ser Gln Leu Glu Lys Leu Ile Glu
                560                 565                 570
Lys Asn Tyr Phe Leu His Lys Ser Ala Gln Glu Ala Tyr Lys Ser
                575                 580                 585
Tyr Ile Arg Ala Tyr Asp Ser His Ser Leu Lys Gln Ile Phe Asn
                590                 595                 600
Val Asn Asn Leu Asn Leu Pro Gln Val Ala Leu Ser Phe Gly Phe
                605                 610                 615
Lys Val Pro Pro Phe Val Asp Leu Asn Val Asn Ser Asn Glu Gly
                620                 625                 630
Lys Gln Lys Lys Arg Gly Gly Gly Gly Gly Phe Gly Tyr Gln Lys
                635                 640                 645
Thr Lys Lys Val Glu Lys Ser Lys Ile Phe Lys His Ile Ser Lys
                650                 655                 660
Lys Ser Ser Asp Ser Arg Gln Phe Ser His
                665                 670
```

<210> 11
<211> 452
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte Identification No.: 1444908CD1

<400> 11

```
Met Glu Phe Gln Ala Val Val Met Ala Val Gly Gly Gly Ser Arg
  1               5                  10                  15
Met Thr Asp Leu Thr Ser Ser Ile Pro Lys Pro Leu Leu Pro Val
             20                  25                  30
Gly Asn Lys Pro Leu Ile Trp Tyr Pro Leu Asn Leu Leu Glu Arg
             35                  40                  45
Val Gly Phe Glu Glu Val Ile Val Val Thr Thr Arg Asp Val Gln
             50                  55                  60
Lys Ala Leu Cys Ala Glu Phe Lys Met Lys Met Lys Pro Asp Ile
             65                  70                  75
Val Cys Ile Pro Asp Asp Ala Asp Met Gly Thr Ala Asp Ser Leu
             80                  85                  90
Arg Tyr Ile Tyr Pro Lys Leu Lys Thr Asp Val Leu Val Leu Ser
             95                 100                 105
Cys Asp Leu Ile Thr Asp Val Ala Leu His Glu Val Val Asp Leu
            110                 115                 120
Phe Arg Ala Tyr Asp Ala Ser Leu Ala Met Leu Met Arg Lys Gly
            125                 130                 135
Gln Asp Ser Ile Glu Pro Val Pro Gly Gln Lys Gly Lys Lys Lys
            140                 145                 150
Ala Val Glu Gln Arg Asp Phe Ile Gly Val Asp Ser Thr Gly Lys
            155                 160                 165
Arg Leu Leu Phe Met Ala Asn Glu Ala Asp Leu Asp Glu Glu Leu
            170                 175                 180
Val Ile Lys Gly Ser Ile Leu Gln Lys His Pro Arg Ile Arg Phe
            185                 190                 195
His Thr Gly Leu Val Asp Ala His Leu Tyr Cys Leu Lys Lys Tyr
            200                 205                 210
Ile Val Asp Phe Leu Met Glu Asn Gly Ser Ile Thr Ser Ile Arg
            215                 220                 225
Ser Glu Leu Ile Pro Tyr Leu Val Arg Lys Gln Phe Ser Ser Ala
            230                 235                 240
Ser Ser Gln Gln Gly Gln Glu Glu Lys Glu Glu Asp Leu Lys Lys
            245                 250                 255
Lys Glu Leu Lys Ser Leu Asp Ile Tyr Ser Phe Ile Lys Glu Ala
            260                 265                 270
Asn Thr Leu Asn Leu Ala Pro Tyr Asp Ala Cys Trp Asn Ala Cys
            275                 280                 285
Arg Gly Asp Arg Trp Glu Asp Leu Ser Arg Ser Gln Val Arg Cys
            290                 295                 300
Tyr Val His Ile Met Lys Glu Gly Leu Cys Ser Arg Val Ser Thr
            305                 310                 315
Leu Gly Leu Tyr Met Glu Ala Asn Arg Gln Val Pro Lys Leu Leu
            320                 325                 330
Ser Ala Leu Cys Pro Glu Glu Pro Pro Val His Ser Ser Ala Gln
            335                 340                 345
Ile Val Ser Lys His Leu Val Gly Val Asp Ser Leu Ile Gly Pro
            350                 355                 360
Glu Thr Gln Ile Gly Glu Lys Ser Ser Ile Lys Arg Ser Val Ile
            365                 370                 375
Gly Ser Ser Cys Leu Ile Lys Asp Arg Val Thr Ile Thr Asn Cys
            380                 385                 390
Leu Leu Met Asn Ser Val Thr Val Glu Glu Gly Ser Asn Ile Gln
            395                 400                 405
```

```
Gly Ser Val Ile Cys Asn Asn Ala Val Ile Glu Lys Gly Ala Asp
                410                 415                 420
Ile Lys Asp Cys Leu Ile Gly Ser Gly Gln Arg Ile Glu Ala Lys
                425                 430                 435
Ala Lys Arg Val Asn Glu Val Ile Val Gly Asn Asp Gln Leu Met
                440                 445                 450
Glu Ile
```

<210> 12
<211> 748
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1557481CD1

<400> 12

```
Met Ala Asp Ser Ser Gly Gln Gln Gly Lys Gly Arg Arg Val Gln
  1                 5                  10                  15
Pro Gln Trp Ser Pro Pro Ala Gly Thr Gln Pro Cys Arg Leu His
                 20                  25                  30
Leu Tyr Asn Ser Leu Thr Arg Asn Lys Glu Val Phe Ile Pro Gln
                 35                  40                  45
Asp Gly Lys Lys Val Thr Trp Tyr Cys Cys Gly Pro Thr Val Tyr
                 50                  55                  60
Asp Ala Ser His Met Gly His Ala Arg Ser Tyr Ile Ser Phe Asp
                 65                  70                  75
Ile Leu Arg Arg Val Leu Lys Asp Tyr Phe Lys Phe Asp Val Phe
                 80                  85                  90
Tyr Cys Met Asn Ile Thr Asp Ile Asp Asp Lys Ile Ile Lys Arg
                 95                 100                 105
Ala Arg Gln Asn His Leu Phe Glu Gln Tyr Arg Glu Lys Arg Pro
                110                 115                 120
Glu Ala Ala Gln Leu Leu Glu Asp Val Gln Ala Ala Leu Lys Pro
                125                 130                 135
Phe Ser Val Lys Leu Asn Glu Thr Thr Asp Pro Asp Lys Lys Gln
                140                 145                 150
Met Leu Glu Arg Ile Gln His Ala Val Gln Leu Ala Thr Glu Pro
                155                 160                 165
Leu Glu Lys Ala Val Gln Ser Arg Leu Thr Gly Glu Glu Val Asn
                170                 175                 180
Ser Cys Val Glu Val Leu Leu Glu Glu Ala Lys Asp Leu Leu Ser
                185                 190                 195
Asp Trp Leu Asp Ser Thr Leu Gly Cys Asp Val Thr Asp Asn Ser
                200                 205                 210
Ile Phe Ser Lys Leu Pro Lys Phe Trp Glu Gly Asp Phe His Arg
                215                 220                 225
Asp Met Glu Ala Leu Asn Val Leu Pro Pro Asp Val Leu Thr Arg
                230                 235                 240
Val Ser Glu Tyr Val Pro Glu Ile Val Asn Phe Val Gln Lys Ile
                245                 250                 255
Val Asp Asn Gly Tyr Gly Tyr Val Ser Asn Gly Ser Val Tyr Phe
                260                 265                 270
```

```
Asp Thr Ala Lys Phe Ala Ser Ser Glu Lys His Ser Tyr Gly Lys
                275                 280                 285
Leu Val Pro Glu Ala Val Gly Asp Gln Lys Ala Leu Gln Glu Gly
                290                 295                 300
Glu Gly Asp Leu Ser Ile Ser Ala Asp Arg Leu Ser Glu Lys Arg
                305                 310                 315
Ser Pro Asn Asp Phe Ala Leu Trp Lys Ala Ser Lys Pro Gly Glu
                320                 325                 330
Pro Ser Trp Pro Cys Pro Trp Gly Lys Gly Arg Pro Gly Trp His
                335                 340                 345
Ile Glu Cys Ser Ala Met Ala Gly Thr Leu Leu Gly Ala Ser Met
                350                 355                 360
Asp Ile His Gly Gly Gly Phe Asp Leu Arg Phe Pro His His Asp
                365                 370                 375
Asn Glu Leu Ala Gln Ser Glu Ala Tyr Phe Glu Asn Asp Cys Trp
                380                 385                 390
Val Arg Tyr Phe Leu His Thr Gly His Leu Thr Ile Ala Gly Cys
                395                 400                 405
Lys Met Ser Lys Ser Leu Lys Asn Phe Ile Thr Ile Lys Asp Ala
                410                 415                 420
Leu Lys Lys His Ser Ala Arg Gln Leu Arg Leu Ala Phe Leu Met
                425                 430                 435
His Ser Trp Lys Asp Thr Leu Asp Tyr Ser Ser Asn Thr Met Glu
                440                 445                 450
Ser Ala Leu Gln Tyr Glu Lys Phe Leu Asn Glu Phe Phe Leu Asn
                455                 460                 465
Val Lys Asp Ile Leu Arg Ala Pro Val Asp Ile Thr Gly Gln Phe
                470                 475                 480
Glu Lys Trp Gly Glu Glu Glu Ala Glu Leu Asn Lys Asn Phe Tyr
                485                 490                 495
Asp Lys Lys Thr Ala Ile His Lys Ala Leu Cys Asp Asn Val Asp
                500                 505                 510
Thr Arg Thr Val Met Glu Glu Met Arg Ala Leu Val Ser Gln Cys
                515                 520                 525
Asn Leu Tyr Met Ala Ala Arg Lys Ala Val Arg Lys Arg Pro Asn
                530                 535                 540
Gln Ala Leu Leu Glu Asn Ile Ala Leu Tyr Leu Thr His Met Leu
                545                 550                 555
Lys Ile Phe Gly Ala Val Glu Glu Asp Ser Ser Leu Gly Phe Pro
                560                 565                 570
Val Gly Gly Pro Gly Thr Ser Leu Ser Leu Glu Ala Thr Val Met
                575                 580                 585
Pro Tyr Leu Gln Val Leu Ser Glu Phe Arg Glu Gly Val Arg Lys
                590                 595                 600
Ile Ala Arg Glu Gln Lys Val Pro Glu Ile Leu Gln Leu Ser Asp
                605                 610                 615
Ala Leu Arg Asp Asn Ile Leu Pro Glu Leu Gly Val Arg Phe Glu
                620                 625                 630
Asp His Glu Gly Leu Pro Thr Val Val Lys Leu Val Asp Arg Asn
                635                 640                 645
Thr Leu Leu Lys Glu Arg Glu Glu Lys Arg Arg Val Glu Glu Glu
                650                 655                 660
Lys Arg Lys Lys Lys Glu Glu Ala Ala Arg Arg Lys Gln Glu Gln
                665                 670                 675
Glu Ala Ala Lys Leu Ala Lys Met Lys Ile Pro Pro Ser Glu Met
                680                 685                 690
Phe Leu Ser Glu Thr Asp Lys Tyr Ser Lys Phe Asp Glu Asn Gly
```

```
                695                 700                 705
Leu Pro Thr His Asp Met Glu Gly Lys Glu Leu Ser Lys Gly Gln
                710                 715                 720
Ala Lys Lys Leu Lys Lys Leu Phe Glu Ala Gln Glu Lys Leu Tyr
                725                 730                 735
Lys Glu Tyr Leu Gln Met Ala Gln Asn Gly Ser Phe Gln
                740                 745
```

<210> 13
<211> 328
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1747456CD1

<400> 13

```
Met Glu Ala Asn Gly Ser Gln Gly Thr Ser Gly Ser Ala Asn Asp
  1               5                  10                  15
Ser Gln His Asp Pro Gly Lys Met Phe Ile Gly Gly Leu Ser Trp
                20                  25                  30
Gln Thr Ser Pro Asp Ser Leu Arg Asp Tyr Phe Ser Lys Phe Gly
                35                  40                  45
Glu Ile Arg Glu Cys Met Val Met Arg Asp Pro Thr Thr Lys Arg
                50                  55                  60
Ser Arg Gly Phe Gly Phe Val Thr Phe Ala Asp Pro Ala Ser Val
                65                  70                  75
Asp Lys Val Leu Gly Gln Pro His His Glu Leu Asp Ser Lys Thr
                80                  85                  90
Ile Asp Pro Lys Val Ala Phe Pro Arg Arg Ala Gln Pro Lys Met
                95                 100                 105
Val Thr Arg Thr Lys Lys Ile Phe Val Gly Gly Leu Ser Ala Asn
               110                 115                 120
Thr Val Val Glu Asp Val Lys Gln Tyr Phe Glu Gln Phe Gly Lys
               125                 130                 135
Val Glu Asp Ala Met Leu Met Phe Asp Lys Thr Thr Asn Arg His
               140                 145                 150
Arg Gly Phe Gly Phe Val Thr Phe Glu Asn Glu Asp Val Val Glu
               155                 160                 165
Lys Val Cys Glu Ile His Phe His Glu Ile Asn Asn Lys Met Val
               170                 175                 180
Glu Cys Lys Lys Ala Gln Pro Lys Glu Val Met Phe Pro Pro Gly
               185                 190                 195
Thr Arg Gly Arg Ala Arg Gly Leu Pro Tyr Thr Met Asp Ala Phe
               200                 205                 210
Met Leu Gly Met Gly Met Leu Gly Tyr Pro Asn Phe Val Ala Thr
               215                 220                 225
Tyr Gly Arg Gly Tyr Pro Gly Phe Ala Pro Ser Tyr Gly Tyr Gln
               230                 235                 240
Phe Pro Gly Phe Pro Ala Ala Ala Tyr Gly Pro Val Ala Ala Ala
               245                 250                 255
Ala Val Ala Ala Ala Arg Gly Ser Gly Ser Asn Pro Ala Arg Pro
               260                 265                 270
```

```
Gly Gly Phe Pro Gly Ala Asn Ser Pro Gly Pro Val Ala Asp Leu
                275                 280                 285
Tyr Gly Pro Ala Ser Gln Asp Ser Gly Val Gly Asn Tyr Ile Ser
                290                 295                 300
Ala Ala Ser Pro Gln Pro Gly Ser Gly Phe Gly His Gly Ile Ala
                305                 310                 315
Gly Pro Leu Ile Ala Thr Ala Phe Thr Asn Gly Tyr His
                320                 325
```

<210> 14
<211> 563
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1748626CD1

<400> 14

```
Met Pro Glu Asp Asp Gln Arg Ala Thr Arg Asn Leu Phe Ile Gly
  1               5                  10                  15
Asn Leu Asp His Ser Val Ser Glu Val Glu Leu Arg Arg Ala Phe
                20                  25                  30
Glu Lys Tyr Gly Ile Ile Glu Glu Val Val Ile Lys Arg Pro Ala
                35                  40                  45
Arg Gly Gln Gly Gly Ala Tyr Ala Phe Leu Lys Phe Gln Asn Leu
                50                  55                  60
Asp Met Ala His Arg Ala Lys Val Ala Met Ser Gly Arg Val Ile
                65                  70                  75
Gly Arg Asn Pro Ile Lys Ile Gly Tyr Gly Lys Ala Asn Pro Thr
                80                  85                  90
Thr Arg Leu Trp Val Gly Gly Leu Gly Pro Asn Thr Ser Leu Ala
                95                 100                 105
Ala Leu Ala Arg Glu Phe Asp Arg Phe Gly Ser Ile Arg Thr Ile
               110                 115                 120
Asp His Val Lys Gly Asp Ser Phe Ala Tyr Ile Gln Tyr Glu Ser
               125                 130                 135
Leu Asp Ala Ala Gln Ala Ala Cys Ala Lys Met Arg Gly Phe Pro
               140                 145                 150
Leu Gly Gly Pro Asp Arg Arg Leu Arg Val Asp Phe Ala Lys Ala
               155                 160                 165
Glu Glu Thr Arg Tyr Pro Gln Gln Tyr Gln Pro Ser Pro Leu Pro
               170                 175                 180
Val His Tyr Glu Leu Leu Thr Asp Gly Tyr Thr Arg His Arg Asn
               185                 190                 195
Leu Asp Ala Asp Leu Val Arg Asp Arg Thr Pro Pro His Leu Leu
               200                 205                 210
Tyr Ser Asp Arg Asp Arg Thr Phe Leu Glu Gly Asp Trp Thr Ser
               215                 220                 225
Pro Ser Lys Ser Ser Asp Arg Arg Asn Ser Leu Glu Gly Tyr Ser
               230                 235                 240
Arg Ser Val Arg Ser Arg Ser Gly Glu Arg Trp Gly Ala Asp Gly
               245                 250                 255
Asp Arg Gly Leu Pro Lys Pro Trp Glu Glu Arg Arg Lys Arg Arg
```

```
                260                 265                 270
Ser Leu Ser Ser Asp Arg Gly Arg Thr Thr His Ser Pro Tyr Glu
                275                 280                 285
Glu Arg Ser Arg Thr Lys Gly Ser Gly Gln Gln Ser Glu Arg Gly
                290                 295                 300
Ser Asp Arg Thr Pro Glu Arg Ser Arg Lys Glu Asn His Ser Ser
                305                 310                 315
Glu Gly Thr Lys Glu Ser Ser Ser Asn Ser Leu Ser Asn Ser Arg
                320                 325                 330
His Gly Ala Glu Glu Arg Gly His His His His His His Glu Ala
                335                 340                 345
Ala Asp Ser Ser His Gly Lys Lys Ala Arg Asp Ser Glu Arg Asn
                350                 355                 360
His Arg Thr Thr Glu Ala Glu Pro Lys Pro Leu Glu Glu Pro Lys
                365                 370                 375
His Glu Thr Lys Lys Leu Lys Asn Leu Ser Glu Tyr Ala Gln Thr
                380                 385                 390
Leu Gln Leu Gly Trp Asn Gly Leu Leu Val Leu Lys Asn Ser Cys
                395                 400                 405
Phe Pro Thr Ser Met His Ile Leu Glu Gly Asp Gln Gly Val Ile
                410                 415                 420
Ser Ser Leu Leu Lys Asp His Thr Ser Gly Ser Lys Leu Thr Gln
                425                 430                 435
Leu Lys Ile Ala Gln Arg Leu Arg Leu Asp Gln Pro Lys Leu Asp
                440                 445                 450
Glu Val Thr Arg Arg Ile Lys Gln Gly Ser Pro Asn Gly Tyr Ala
                455                 460                 465
Val Leu Leu Ala Thr Gln Ala Thr Pro Ser Gly Leu Gly Thr Glu
                470                 475                 480
Gly Met Pro Thr Val Glu Pro Gly Leu Gln Arg Arg Leu Leu Arg
                485                 490                 495
Asn Leu Val Ser Tyr Leu Lys Gln Lys Gln Ala Ala Gly Val Ile
                500                 505                 510
Ser Leu Pro Val Gly Gly Ser Lys Gly Arg Asp Gly Thr Gly Met
                515                 520                 525
Leu Tyr Ala Phe Pro Pro Cys Asp Phe Ser Gln Gln Tyr Leu Gln
                530                 535                 540
Ser Ala Leu Arg Thr Leu Gly Lys Leu Glu Glu Glu His Met Val
                545                 550                 555
Ile Val Ile Val Arg Asp Thr Ala
                560
```

<210> 15
<211> 153
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1879135CD1

<400> 15

```
Met Met Ser Gln Ser Gly His Glu Tyr Asp Pro Ile Asn Tyr Met
  1               5                  10                  15
```

```
Lys Lys Pro Leu Gly Pro Pro Pro Pro Ser Tyr Thr Cys Phe Arg
                  20                  25                  30
Cys Gly Lys Pro Gly His Tyr Ile Lys Asn Cys Pro Thr Asn Gly
                  35                  40                  45
Asp Lys Asn Phe Glu Ser Gly Pro Arg Ile Lys Lys Ser Thr Gly
                  50                  55                  60
Ile Pro Arg Ser Phe Met Met Glu Val Lys Asp Pro Asn Met Lys
                  65                  70                  75
Gly Ala Met Leu Thr Asn Thr Gly Lys Tyr Ala Ile Pro Thr Ile
                  80                  85                  90
Asp Ala Glu Ala Tyr Ala Ile Gly Lys Lys Glu Lys Pro Pro Phe
                  95                 100                 105
Leu Pro Glu Glu Pro Ser Ser Ser Ser Glu Glu Asp Asp Pro Ile
                 110                 115                 120
Pro Asp Glu Leu Leu Cys Leu Ile Cys Lys Asp Ile Met Thr Asp
                 125                 130                 135
Ala Val Val Ile Pro Cys Cys Gly Asn Ser Tyr Cys Asp Glu Cys
                 140                 145                 150
Lys Lys Cys
```

<210> 16
<211> 286
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2073417CD1

<400> 16

```
Met Ser Trp Leu Leu Phe Leu Ala His Arg Val Ala Leu Ala Ala
  1               5                  10                  15
Leu Pro Cys Arg Arg Gly Ser Arg Gly Phe Gly Met Phe Tyr Ala
                  20                  25                  30
Val Arg Arg Gly Arg Lys Thr Gly Val Phe Leu Thr Trp Asn Glu
                  35                  40                  45
Cys Arg Ala Gln Val Asp Arg Phe Pro Ala Ala Arg Phe Lys Lys
                  50                  55                  60
Phe Ala Thr Glu Asp Glu Ala Trp Ala Phe Val Arg Lys Ser Ala
                  65                  70                  75
Ser Pro Glu Val Ser Glu Gly His Glu Asn Gln His Gly Gln Glu
                  80                  85                  90
Ser Glu Ala Lys Ala Ser Lys Arg Leu Arg Glu Pro Leu Asp Gly
                  95                 100                 105
Asp Gly His Glu Ser Ala Glu Pro Tyr Ala Lys His Met Lys Pro
                 110                 115                 120
Ser Met Glu Pro Ala Pro Pro Val Ser Arg Asp Thr Phe Ser Tyr
                 125                 130                 135
Met Gly Asp Phe Val Val Val Tyr Thr Asp Gly Cys Cys Ser Ser
                 140                 145                 150
Asn Gly Arg Arg Arg Pro Arg Ala Gly Ile Gly Val Tyr Trp Gly
                 155                 160                 165
Pro Gly His Pro Leu Asn Val Gly Ile Arg Leu Pro Gly Arg Gln
                 170                 175                 180
```

```
Thr Asn Gln Arg Ala Glu Ile His Ala Ala Cys Lys Ala Ile Glu
                185                 190                 195
Gln Ala Lys Thr Gln Asn Ile Asn Lys Leu Val Leu Tyr Thr Asp
                200                 205                 210
Ser Met Phe Thr Ile Asn Gly Ile Thr Asn Trp Val Gln Gly Trp
                215                 220                 225
Lys Lys Asn Gly Trp Lys Thr Ser Ala Gly Lys Glu Val Ile Asn
                230                 235                 240
Lys Glu Asp Phe Val Ala Leu Glu Arg Leu Thr Gln Gly Met Asp
                245                 250                 255
Ile Gln Trp Met His Val Pro Gly His Ser Gly Phe Ile Gly Asn
                260                 265                 270
Glu Glu Ala Asp Arg Leu Ala Arg Glu Gly Ala Lys Gln Ser Glu
                275                 280                 285
Asp
```

<210> 17
<211> 537
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2129080CD1

<400> 17

```
Met Leu Ala Arg Glu Thr Tyr Glu Glu Asp Arg Glu Tyr Glu Ser
  1               5                  10                  15
Gln Ala Lys Arg Leu Lys Thr Glu Glu Gly Glu Ile Asp Tyr Ser
                20                  25                  30
Ala Glu Glu Gly Glu Asn Arg Arg Glu Ala Thr Pro Arg Gly Gly
                35                  40                  45
Gly Asp Gly Gly Gly Gly Gly Arg Ser Phe Ser Gln Pro Glu Ala
                50                  55                  60
Gly Gly Ser His His Lys Val Ser Val Ser Pro Val Val His Val
                65                  70                  75
Arg Gly Leu Cys Glu Ser Val Val Glu Ala Asp Leu Val Glu Ala
                80                  85                  90
Leu Glu Lys Phe Gly Thr Ile Cys Tyr Val Met Met Met Pro Phe
                95                 100                 105
Lys Arg Gln Ala Leu Val Glu Phe Glu Asn Ile Asp Ser Ala Lys
                110                 115                 120
Glu Cys Val Thr Phe Ala Ala Asp Glu Pro Val Tyr Ile Ala Gly
                125                 130                 135
Gln Gln Ala Phe Phe Asn Tyr Ser Thr Ser Lys Arg Ile Thr Arg
                140                 145                 150
Pro Gly Asn Thr Asp Asp Pro Ser Gly Gly Asn Lys Val Leu Leu
                155                 160                 165
Leu Ser Ile Gln Asn Pro Leu Tyr Pro Ile Thr Val Asp Val Leu
                170                 175                 180
Tyr Thr Val Cys Asn Pro Val Gly Lys Val Gln Arg Ile Val Ile
                185                 190                 195
Phe Lys Arg Asn Gly Ile Gln Ala Met Val Glu Phe Glu Ser Val
                200                 205                 210
```

```
Leu Cys Ala Gln Lys Ala Lys Ala Ala Leu Asn Gly Ala Asp Ile
                215                 220                 225
Tyr Ala Gly Cys Cys Thr Leu Lys Ile Glu Tyr Ala Arg Pro Thr
                230                 235                 240
Arg Leu Asn Val Ile Arg Asn Asp Asn Asp Ser Trp Asp Tyr Thr
                245                 250                 255
Lys Pro Tyr Leu Gly Arg Arg Asp Arg Gly Lys Gly Arg Gln Arg
                260                 265                 270
Gln Ala Ile Leu Gly Glu His Pro Ser Ser Phe Arg His Asp Gly
                275                 280                 285
Tyr Gly Ser His Gly Pro Leu Leu Pro Leu Pro Ser Arg Tyr Arg
                290                 295                 300
Met Gly Ser Arg Asp Thr Pro Glu Leu Val Ala Tyr Pro Leu Pro
                305                 310                 315
Gln Ala Ser Ser Ser Tyr Met His Gly Gly Asn Pro Ser Gly Ser
                320                 325                 330
Val Val Met Val Ser Gly Leu His Gln Leu Lys Met Asn Cys Ser
                335                 340                 345
Arg Val Phe Asn Leu Phe Cys Leu Tyr Gly Asn Ile Glu Lys Val
                350                 355                 360
Lys Phe Met Lys Thr Ile Pro Gly Thr Ala Leu Val Glu Met Gly
                365                 370                 375
Asp Glu Tyr Ala Val Glu Arg Ala Val Thr His Leu Asn Asn Val
                380                 385                 390
Lys Leu Phe Gly Lys Arg Leu Asn Val Cys Val Ser Lys Gln His
                395                 400                 405
Ser Val Val Pro Ser Gln Ile Phe Glu Leu Glu Asp Gly Thr Ser
                410                 415                 420
Ser Tyr Lys Asp Phe Ala Met Ser Lys Asn Asn Arg Phe Thr Ser
                425                 430                 435
Ala Gly Gln Ala Ser Lys Asn Ile Ile Gln Pro Pro Ser Cys Val
                440                 445                 450
Leu His Tyr Tyr Asn Val Pro Leu Cys Val Thr Glu Glu Thr Phe
                455                 460                 465
Thr Lys Leu Cys Asn Asp His Glu Val Leu Thr Phe Ile Lys Tyr
                470                 475                 480
Lys Val Phe Asp Ala Lys Pro Ser Ala Lys Thr Leu Ser Gly Leu
                485                 490                 495
Leu Glu Trp Glu Cys Lys Thr Asp Ala Val Glu Ala Leu Thr Ala
                500                 505                 510
Leu Asn His Tyr Gln Ile Arg Val Pro Asn Gly Ser Asn Pro Tyr
                515                 520                 525
Thr Leu Lys Leu Cys Phe Ser Thr Ser Ser His Leu
                530                 535
```

<210> 18
<211> 163
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2472867CD1

<400> 18
Met Arg Ile Glu Lys Cys Tyr Phe Cys Ser Gly Pro Ile Tyr Pro
1 5 10 15
Gly His Gly Met Met Phe Val Arg Asn Asp Cys Lys Val Phe Arg
20 25 30
Phe Cys Lys Ser Lys Cys His Lys Asn Phe Lys Lys Lys Arg Asn
35 40 45
Pro Arg Lys Val Arg Trp Thr Lys Ala Phe Arg Lys Ala Ala Gly
50 55 60
Lys Glu Leu Thr Val Asp Asn Ser Phe Glu Phe Glu Lys Arg Arg
65 70 75
Asn Glu Pro Ile Lys Tyr Gln Arg Glu Leu Trp Asn Lys Thr Ile
80 85 90
Asp Ala Met Lys Arg Val Glu Glu Ile Lys Gln Lys Arg Gln Ala
95 100 105
Lys Phe Ile Met Asn Arg Leu Lys Lys Asn Lys Glu Leu Gln Lys
110 115 120
Val Gln Asp Ile Lys Glu Val Lys Gln Asn Ile His Leu Ile Arg
125 130 135
Ala Pro Leu Ala Gly Lys Gly Lys Gln Leu Glu Glu Lys Met Val
140 145 150
Gln Gln Leu Gln Glu Asp Val Asp Met Glu Asp Ala Pro
155 160

<210> 19
<211> 178
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2764755CD1

<400> 19
Met Ser Ser Phe Ser Arg Ala Pro Gln Gln Trp Ala Thr Phe Ala
1 5 10 15
Arg Ile Trp Tyr Leu Leu Asp Gly Lys Met Gln Pro Pro Gly Lys
20 25 30
Leu Ala Ala Met Ala Ser Ile Arg Leu Gln Gly Leu His Lys Pro
35 40 45
Val Tyr His Ala Leu Ser Asp Cys Gly Asp His Val Val Ile Met
50 55 60
Asn Thr Arg His Ile Ala Phe Ser Gly Asn Lys Trp Glu Gln Lys
65 70 75
Val Tyr Ser Ser His Thr Gly Tyr Pro Gly Gly Phe Arg Gln Val
80 85 90
Thr Ala Ala Gln Leu His Leu Arg Asp Pro Val Ala Ile Val Lys
95 100 105
Leu Ala Ile Tyr Gly Met Leu Pro Lys Asn Leu His Arg Arg Thr
110 115 120
Met Met Glu Arg Leu His Leu Phe Pro Asp Glu Tyr Ile Pro Glu
125 130 135
Asp Ile Leu Lys Asn Leu Val Glu Glu Leu Pro Gln Pro Arg Lys
140 145 150

```
Ile Pro Lys Arg Leu Asp Glu Tyr Thr Gln Glu Glu Ile Asp Ala
                155                 160                 165
Phe Pro Arg Leu Trp Thr Pro Pro Glu Asp Tyr Arg Leu
                170                 175
```

<210> 20
<211> 140
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2875939CD1

<400> 20

```
Met Ala Glu Asn Arg Glu Pro Arg Gly Ala Val Glu Ala Glu Leu
  1               5                  10                  15
Asp Pro Val Glu Tyr Thr Leu Arg Lys Arg Leu Pro Ser Arg Leu
                20                  25                  30
Pro Arg Arg Pro Asn Asp Ile Tyr Val Asn Met Lys Thr Asp Phe
                35                  40                  45
Lys Ala Gln Leu Ala Arg Cys Gln Lys Leu Leu Asp Gly Gly Ala
                50                  55                  60
Arg Gly Gln Asn Ala Cys Ser Glu Ile Tyr Ile His Gly Leu Gly
                65                  70                  75
Leu Ala Ile Asn Arg Ala Ile Asn Ile Ala Leu Gln Leu Gln Ala
                80                  85                  90
Gly Ser Phe Gly Ser Leu Gln Val Ala Ala Asn Thr Ser Thr Val
                95                 100                 105
Glu Leu Val Asp Glu Leu Glu Pro Glu Thr Asp Thr Arg Glu Pro
               110                 115                 120
Leu Thr Arg Ile Arg Asn Asn Ser Ala Ile His Ile Arg Val Phe
               125                 130                 135
Arg Val Thr Pro Lys
               140
```

<210> 21
<211> 209
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 3591363CD1

<400> 21

```
Met Ala Ala Ala Met Ala Ala Ser Ser Leu Thr Val Thr Leu Gly
  1               5                  10                  15
Arg Leu Ala Ser Ala Cys Ser His Ser Ile Leu Arg Pro Ser Gly
                20                  25                  30
Pro Gly Ala Ala Ser Leu Trp Ser Ala Ser Arg Arg Phe Asn Ser
```

```
                  35                  40                  45
Gln Ser Thr Ser Tyr Leu Pro Gly Tyr Val Pro Lys Thr Ser Leu
                  50                  55                  60
Ser Ser Pro Pro Trp Pro Glu Val Val Leu Pro Asp Pro Val Glu
                  65                  70                  75
Glu Thr Arg His His Ala Glu Val Val Lys Lys Val Asn Glu Met
                  80                  85                  90
Ile Val Thr Gly Gln Tyr Gly Arg Leu Phe Ala Val Val His Phe
                  95                 100                 105
Ala Ser Arg Gln Trp Lys Val Thr Ser Glu Asp Leu Ile Leu Ile
                 110                 115                 120
Gly Asn Glu Leu Asp Leu Ala Cys Gly Glu Arg Ile Arg Leu Glu
                 125                 130                 135
Lys Val Leu Leu Val Gly Ala Asp Asn Phe Thr Leu Leu Gly Lys
                 140                 145                 150
Pro Leu Leu Gly Lys Asp Leu Val Arg Val Glu Ala Thr Val Ile
                 155                 160                 165
Glu Lys Thr Glu Ser Trp Pro Arg Ile Ile Met Arg Phe Arg Lys
                 170                 175                 180
Arg Lys Asn Phe Lys Lys Lys Arg Ile Val Thr Thr Pro Gln Thr
                 185                 190                 195
Val Leu Arg Ile Asn Ser Ile Glu Ile Ala Pro Cys Leu Leu
                 200                 205
```

<210> 22
<211> 162
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 3702292CD1

<400> 22

```
Met Ala Pro Lys Ala Lys Glu Ala Pro Ala His Pro Lys Ala Glu
  1                 5                  10                  15
Ala Lys Ala Lys Ala Leu Lys Ala Lys Lys Ala Val Leu Lys Gly
                  20                  25                  30
Val Arg Ser His Thr Gln Lys Gln Lys Ile Arg Met Ser Leu Thr
                  35                  40                  45
Phe Arg Arg Pro Lys Thr Leu Arg Leu Arg Arg Gln Pro Arg Tyr
                  50                  55                  60
Pro Arg Lys Ser Thr Pro Arg Arg Asn Lys Leu Gly His Tyr Ala
                  65                  70                  75
Ile Ile Lys Phe Pro Leu Ala Thr Glu Ser Ala Val Lys Lys Ile
                  80                  85                  90
Glu Glu Asn Asn Thr Leu Val Phe Thr Val Asp Val Lys Ala Asn
                  95                 100                 105
Lys His Gln Ile Arg Gln Ala Val Lys Lys Leu Tyr Asp Ser Asp
                 110                 115                 120
Val Ala Lys Val Thr Thr Leu Ile Cys Pro Asp Lys Glu Asn Lys
                 125                 130                 135
Ala Tyr Val Arg Leu Ala Pro Asp Tyr Asp Ala Phe Asp Val Val
                 140                 145                 150
```

```
Thr Lys Leu Gly Ser Pro Lys Leu Ser Pro Ala Gly
                155                 160
```

<210> 23
<211> 623
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 3778908CD1

<400> 23

```
Met Ala Thr Glu His Val Asn Gly Asn Gly Thr Glu Glu Pro Met
  1               5                  10                  15
Asp Thr Thr Ser Ala Val Ile His Ser Glu Asn Phe Gln Thr Leu
             20                  25                  30
Leu Asp Ala Gly Leu Pro Gln Lys Val Ala Glu Lys Leu Asp Glu
             35                  40                  45
Ile Tyr Val Ala Gly Leu Val Ala His Ser Asp Leu Asp Glu Arg
             50                  55                  60
Ala Ile Glu Ala Leu Lys Glu Phe Asn Glu Asp Gly Ala Leu Ala
             65                  70                  75
Val Leu Gln Gln Phe Lys Asp Ser Asp Leu Ser His Val Gln Asn
             80                  85                  90
Lys Ser Ala Phe Leu Cys Gly Val Met Lys Thr Tyr Arg Gln Arg
             95                 100                 105
Glu Lys Gln Gly Thr Lys Val Ala Asp Ser Ser Lys Gly Pro Asp
            110                 115                 120
Glu Ala Lys Ile Lys Ala Leu Leu Glu Arg Thr Gly Tyr Thr Leu
            125                 130                 135
Asp Val Thr Thr Gly Gln Arg Lys Tyr Gly Gly Pro Pro Pro Asp
            140                 145                 150
Ser Val Tyr Ser Gly Gln Gln Pro Ser Val Gly Thr Glu Ile Phe
            155                 160                 165
Val Gly Lys Ile Pro Arg Asp Leu Phe Glu Asp Glu Leu Val Pro
            170                 175                 180
Leu Phe Glu Lys Ala Gly Pro Ile Trp Asp Leu Arg Leu Met Met
            185                 190                 195
Asp Pro Leu Thr Gly Leu Asn Arg Gly Tyr Ala Phe Val Thr Phe
            200                 205                 210
Cys Thr Lys Glu Ala Ala Gln Glu Ala Val Lys Leu Tyr Asn Asn
            215                 220                 225
His Glu Ile Arg Ser Gly Lys His Ile Gly Val Cys Ile Ser Val
            230                 235                 240
Ala Asn Asn Arg Leu Phe Val Gly Ser Ile Pro Lys Ser Lys Thr
            245                 250                 255
Lys Glu Gln Ile Leu Glu Glu Phe Ser Lys Val Thr Glu Gly Leu
            260                 265                 270
Thr Asp Val Ile Leu Tyr His Gln Pro Asp Asp Lys Lys Lys Asn
            275                 280                 285
Arg Gly Phe Cys Phe Leu Glu Tyr Glu Asp His Lys Thr Ala Ala
            290                 295                 300
Gln Ala Arg Arg Arg Leu Met Ser Gly Lys Val Lys Val Trp Gly
```

```
                305                 310                 315
Asn Val Gly Thr Val Glu Trp Ala Asp Pro Ile Glu Asp Pro Asp
                320                 325                 330
Pro Glu Val Met Ala Lys Val Lys Val Leu Phe Val Arg Asn Leu
                335                 340                 345
Ala Asn Thr Val Thr Glu Glu Ile Leu Glu Lys Ala Phe Ser Gln
                350                 355                 360
Phe Gly Lys Leu Glu Arg Val Lys Lys Leu Lys Asp Tyr Ala Phe
                365                 370                 375
Ile His Phe Asp Glu Arg Asp Gly Ala Val Lys Ala Met Glu Glu
                380                 385                 390
Met Asn Gly Lys Asp Leu Glu Gly Glu Asn Ile Glu Ile Val Phe
                395                 400                 405
Ala Lys Pro Pro Asp Gln Lys Arg Lys Glu Arg Lys Ala Gln Arg
                410                 415                 420
Gln Ala Ala Lys Asn Gln Met Tyr Asp Asp Tyr Tyr Tyr Tyr Gly
                425                 430                 435
Pro Pro His Met Pro Pro Pro Thr Arg Gly Arg Gly Arg Gly Gly
                440                 445                 450
Arg Gly Gly Tyr Gly Tyr Pro Pro Asp Tyr Tyr Gly Tyr Glu Asp
                455                 460                 465
Tyr Tyr Asp Tyr Tyr Gly Tyr Asp Tyr His Asn Tyr Arg Gly Gly
                470                 475                 480
Tyr Glu Asp Pro Tyr Tyr Gly Tyr Glu Asp Phe Gln Val Gly Ala
                485                 490                 495
Arg Gly Arg Gly Gly Arg Gly Ala Arg Gly Ala Ala Pro Ser Arg
                500                 505                 510
Gly Arg Gly Ala Ala Pro Pro Arg Gly Arg Ala Gly Tyr Ser Gln
                515                 520                 525
Arg Gly Gly Pro Gly Ser Ala Arg Gly Val Arg Gly Ala Arg Gly
                530                 535                 540
Gly Ala Gln Gln Gln Arg Gly Arg Gly Val Arg Gly Ala Arg Gly
                545                 550                 555
Gly Arg Gly Gly Asn Val Gly Gly Lys Arg Lys Ala Asp Gly Tyr
                560                 565                 570
Asn Gln Pro Asp Ser Lys Arg Arg Gln Thr Asn Asn Gln Asn Trp
                575                 580                 585
Gly Ser Gln Pro Ile Ala Gln Gln Pro Leu Gln Gly Gly Asp His
                590                 595                 600
Ser Gly Asn Tyr Gly Tyr Lys Ser Glu Asn Gln Glu Phe Tyr Gln
                605                 610                 615
Asp Thr Phe Gly Gln Gln Trp Lys
                620
```

<210> 24
<211> 786
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 4163642CD1

<400> 24

```
Met Ser Phe Ser Arg Ala Leu Leu Trp Ala Arg Leu Pro Ala Gly
 1               5                  10                  15
Arg Gln Ala Gly His Arg Ala Ala Ile Cys Ser Ala Leu Arg Pro
            20                  25                  30
His Phe Gly Pro Phe Pro Gly Val Leu Gly Gln Val Ser Val Leu
        35                  40                  45
Ala Thr Ala Ser Ser Ser Ala Ser Gly Gly Ser Lys Ile Pro Asn
    50                  55                  60
Thr Ser Leu Phe Val Pro Leu Thr Val Lys Pro Gln Gly Pro Ser
65                  70                  75
Ala Asp Gly Asp Val Gly Ala Glu Leu Thr Arg Pro Leu Asp Lys
                80                  85                  90
Asn Glu Val Lys Lys Val Leu Asp Lys Phe Tyr Lys Arg Lys Glu
            95                  100                 105
Ile Gln Lys Leu Gly Ala Asp Tyr Gly Leu Asp Ala Arg Leu Phe
        110                 115                 120
His Gln Ala Phe Ile Ser Phe Arg Asn Tyr Ile Met Gln Ser His
    125                 130                 135
Ser Leu Asp Val Asp Ile His Ile Val Leu Asn Asp Ile Cys Phe
140                 145                 150
Gly Ala Ala His Ala Asp Asp Leu Phe Pro Phe Phe Leu Arg His
155                 160                 165
Ala Lys Gln Ile Phe Pro Val Leu Asp Cys Lys Asp Asp Leu Arg
                170                 175                 180
Lys Ile Ser Asp Leu Arg Ile Pro Pro Asn Trp Tyr Pro Asp Ala
            185                 190                 195
Arg Ala Met Gln Arg Lys Ile Ile Phe His Ser Gly Pro Thr Asn
        200                 205                 210
Ser Gly Lys Thr Tyr His Ala Ile Gln Lys Tyr Phe Ser Ala Lys
    215                 220                 225
Ser Gly Val Tyr Cys Gly Pro Leu Lys Leu Leu Ala His Glu Ile
230                 235                 240
Phe Glu Lys Ser Asn Ala Ala Gly Val Pro Cys Asp Leu Val Thr
                245                 250                 255
Gly Glu Glu Arg Val Thr Val Gln Pro Asn Gly Lys Gln Ala Ser
            260                 265                 270
His Val Ser Cys Thr Val Glu Met Cys Ser Val Thr Thr Pro Tyr
        275                 280                 285
Glu Val Ala Val Ile Asp Glu Ile Gln Met Ile Arg Asp Pro Ala
    290                 295                 300
Arg Gly Trp Ala Trp Thr Arg Ala Leu Leu Gly Leu Cys Ala Glu
305                 310                 315
Glu Val His Leu Cys Gly Glu Pro Ala Ala Ile Asp Leu Val Met
                320                 325                 330
Glu Leu Met Tyr Thr Thr Gly Glu Glu Val Glu Val Arg Asp Tyr
            335                 340                 345
Lys Arg Leu Thr Pro Ile Ser Val Leu Asp His Ala Leu Glu Ser
        350                 355                 360
Leu Asp Asn Leu Arg Pro Gly Asp Cys Ile Val Cys Phe Ser Lys
    365                 370                 375
Asn Asp Ile Tyr Ser Val Ser Arg Gln Ile Glu Ile Arg Gly Leu
380                 385                 390
Glu Ser Ala Val Ile Tyr Gly Ser Leu Pro Pro Gly Thr Lys Leu
395                 400                 405
Ala Gln Ala Lys Lys Phe Asn Asp Pro Asn Asp Pro Cys Lys Ile
                410                 415                 420
Leu Val Ala Thr Asp Ala Ile Gly Met Gly Leu Asn Leu Ser Ile
```

```
                425                 430                 435
Arg Arg Ile Ile Phe Tyr Ser Leu Ile Lys Pro Ser Ile Asn Glu
                440                 445                 450
Lys Gly Glu Arg Glu Leu Glu Pro Ile Thr Thr Ser Gln Ala Leu
                455                 460                 465
Gln Ile Ala Gly Arg Ala Gly Arg Phe Ser Ser Arg Phe Lys Glu
                470                 475                 480
Gly Glu Val Thr Thr Met Asn His Glu Asp Leu Ser Leu Leu Lys
                485                 490                 495
Glu Ile Leu Lys Arg Pro Val Asp Pro Ile Arg Ala Ala Gly Leu
                500                 505                 510
His Pro Thr Ala Glu Gln Ile Glu Met Phe Ala Tyr His Leu Pro
                515                 520                 525
Asp Ala Thr Leu Ser Asn Leu Ile Asp Ile Phe Val Asp Phe Ser
                530                 535                 540
Gln Val Asp Gly Gln Tyr Phe Val Cys Asn Met Asp Asp Phe Lys
                545                 550                 555
Phe Ser Ala Glu Leu Ile Gln His Ile Pro Leu Ser Leu Arg Val
                560                 565                 570
Arg Tyr Val Phe Cys Thr Ala Pro Ile Asn Lys Lys Gln Pro Phe
                575                 580                 585
Val Cys Ser Ser Leu Leu Gln Phe Ala Arg Gln Tyr Ser Arg Asn
                590                 595                 600
Glu Pro Leu Thr Phe Ala Trp Leu Arg Arg Tyr Ile Lys Trp Pro
                605                 610                 615
Leu Leu Pro Pro Lys Asn Ile Lys Asp Leu Met Asp Leu Glu Ala
                620                 625                 630
Val His Asp Val Leu Asp Leu Tyr Leu Trp Leu Ser Tyr Arg Phe
                635                 640                 645
Met Asp Met Phe Pro Asp Ala Ser Leu Ile Arg Asp Leu Gln Lys
                650                 655                 660
Glu Leu Asp Gly Ile Ile Gln Asp Gly Val His Asn Ile Thr Lys
                665                 670                 675
Leu Ile Lys Met Ser Glu Thr His Lys Leu Leu Asn Leu Glu Gly
                680                 685                 690
Phe Pro Ser Gly Ser Gln Ser Arg Leu Ser Gly Thr Leu Lys Ser
                695                 700                 705
Gln Ala Arg Arg Thr Arg Gly Thr Lys Ala Leu Gly Ser Lys Ala
                710                 715                 720
Thr Glu Pro Pro Ser Pro Asp Ala Gly Glu Leu Ser Leu Ala Ser
                725                 730                 735
Arg Leu Val Gln Gln Gly Leu Leu Thr Pro Asp Met Leu Lys Gln
                740                 745                 750
Leu Glu Lys Glu Trp Met Thr Gln Gln Thr Glu His Asn Lys Glu
                755                 760                 765
Lys Thr Glu Ser Gly Thr His Pro Lys Gly Thr Arg Arg Lys Lys
                770                 775                 780
Lys Glu Pro Asp Ser Asp
                785
```

<210> 25
<211> 260
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 4906154CD1

<400> 25

```
Met Thr Pro Val Gln Arg Gly Gly Pro Gly Ala Leu Val Ala Leu
  1               5                  10                  15
Gly Trp Gly Arg Arg Lys Ala Glu Asp Lys Glu Trp Met Pro Val
             20                  25                  30
Thr Lys Leu Gly Arg Leu Val Lys Asp Met Lys Ile Lys Ser Leu
             35                  40                  45
Glu Glu Ile Tyr Leu Phe Ser Leu Pro Ile Lys Glu Ser Glu Ile
             50                  55                  60
Ile Asp Phe Phe Leu Gly Ala Ser Leu Lys Asp Glu Val Leu Lys
             65                  70                  75
Ile Met Pro Val Gln Lys Gln Thr Arg Ala Gly Gln Arg Thr Arg
             80                  85                  90
Phe Lys Ala Phe Val Ala Ile Gly Asp Tyr Asn Gly His Val Gly
             95                 100                 105
Leu Gly Val Lys Cys Ser Lys Glu Val Ala Thr Ala Ile Arg Gly
            110                 115                 120
Ala Ile Ile Leu Ala Lys Leu Ser Ile Val Pro Val Arg Arg Gly
            125                 130                 135
Tyr Trp Gly Asn Lys Ile Gly Lys Pro His Thr Val Pro Cys Lys
            140                 145                 150
Val Thr Gly Arg Cys Gly Ser Val Leu Val Arg Leu Ile Pro Ala
            155                 160                 165
Pro Arg Gly Thr Gly Ile Val Ser Ala Pro Val Pro Lys Lys Leu
            170                 175                 180
Leu Met Met Ala Gly Ile Asp Asp Cys Tyr Thr Ser Ala Arg Gly
            185                 190                 195
Cys Thr Ala Thr Leu Gly Asn Phe Ala Lys Ala Thr Phe Asp Ala
            200                 205                 210
Ile Ser Lys Thr Tyr Ser Tyr Leu Thr Pro Asp Leu Trp Lys Glu
            215                 220                 225
Thr Val Phe Thr Lys Ser Pro Tyr Gln Glu Phe Thr Asp His Leu
            230                 235                 240
Val Lys Thr His Thr Arg Val Ser Val Gln Arg Thr Gln Ala Pro
            245                 250                 255
Ala Val Ala Thr Thr
            260
```

<210> 26
<211> 1872
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 399781CB1

<400> 26

```
gccctctagc tgtgtgtgtc tgaggctcgg ccgcctgagc cgcggacggt ttgctgagcc  60
cgttagtgcg cccggccgag acacgccgcc gccatgtccc gctacctgcg tccccccaac 120
```

acgtctctgt tcgtcaggaa cgtggccgac gacaccaggt ctgaagactt gcggcgtgaa 180
tttggtcgtt atggtcctat agttgatgtg tatgttccac ttgatttcta cactcgccgt 240
ccaagaggat ttgcttatgt tcaatttgag gatgttcgtg atgctgaaga cgctttacat 300
aatttggaca gaaagtggat ttgtggacgg cagattgaaa tacagtttgc ccaggggat 360
cgaaagacac caaatcagat gaaagccaag gaagggagga atgtgtacag ttcttcacgc 420
tatgatgatt atgacagata cagacgttct agaagccgaa gttatgaaag gaggagatca 480
agaagtcggt cttttgatta caactataga agatcgtata gtcctagaaa cagtagaccg 540
actggaagac cacggcgtag agaagccatt ccgacaatga tagaccaaac tgcagctgga 600
atacccagta cagttctgct tactacactt caagaaagat ctgaaagcgg aaaaagaacc 660
aaagaagggc agttcaagcg accaaagggt gggtggaagg tgctgcagta tgaatactgt 720
acgaatattt tgactctggt ctgaaaagat aaaagaatgt tatcgaaaac tacatggaat 780
aattgaagtc ccttcaagtt tgaaagtaag cattttagga caaataaaag gaaattcaac 840
tttgtacttg tggaaactaa tccctaaata tgaataggtt tatattgatt catgggtaac 900
aggtccataa taaattattg gaaactagga tgtctgaata tcaaggaaga cagccatagt 960
ctcttacagt gcctctgttg gtctgtctca aactgaattg ggtgggaaaa ggtatggtcc 1020
aatataaaag ttccattttt gccattattg gcaaatcttg cctttgttta ttttggtgcc 1080
agtgttttct gcttaatcat ttgctttgtt ggcatctgtg tttatttact tgtacaccac 1140
atgcagttta catctgtctt aactactcct tcccaggtaa attccaatta tatttgacat 1200
ccagctaaga gggcccatct cttctcacct ctttcctagt cagtatattc agcaaatatt 1260
tattgagccc ttactgtggg caaatcattg tactggataa ttgagaaaaa tagataattc 1320
ccttattcag taaatgtcta ctgagcacaa tctagtgaat cattacagta tggcctcatt 1380
gttttgtttg aggtgtgtta ttcataacaa tattttacac cattcgtatc aatgtaatta 1440
tagaacacaa tatacgatca aggataagta attgtgtggt tatctgccat ttaaaagtat 1500
ccagtatttg atcacattat tataaataat gaaaaaatga tttaatctgt aataaactgg 1560
tttattgtgc agtgactgta atatactaga gttataataa attgtttact ctgcctcacc 1620
aaacacatgc taggatataa cccccaaaat aagtatttaa ctttgcatta ggtataaagg 1680
agactgggtg ctataattag attattttga ggcagacaga gagctgttat cctaactgat 1740
ttagtatgtt ctgtaattga gaaaatgttc accaaattat acttttttagt gatttacatg 1800
tacattttat aggggacatg ttctgtgtat agcgaataaa taacttttat agtatcaaaa 1860
aaaaaaaaaa aa 1872

<210> 27
<211> 3834
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1806542CB1

<400> 27
ctgcggcggc ggcggcggcg ttaccggccc tcgcgctctt tcccttcctg gggcgccgac 60
cccgccccgct tgcttgcttg cttgcttgcc tgcctgcctg cctgcccggc gccacgcaag 120
agaaggtgcc aggggacgca gagcgactag aggcgcgcgg tcccggccag caccgtctct 180
ggcgttgtag ctgcggccgt ggcggaggac tacggcgaca aggacgaggg ccgctctccc 240
agctctctgc gtgccgcgcc gctccgctcc gctggctgac catctggagt gcaggctggg 300
aggcgggatg gagtgatagg gaagatgttt ataaattctt ctgtgggatc agagggcacg 360
cctattacaa ccagaaaact acaagtataa cagcgaggat ggatgaacag gctctattag 420
ggctaaatcc aaatgctgat tcagacttta gacaaagggc cctggcctat tttgagcagt 480
taaaaatttc cccagatgcc tggcaggtgt gtgcagaagc tctagcccag aggacataca 540
gtgatgatca tgtgaagttt ttctgctttc aagtactgga acatcaagtt aaatacaaat 600
actcagaact aaccactgtt caacaacagc taattaggga gacgctcata tcatggctgc 660
aagctcagat gctgaatccc caaccagaga agacctttat acgaaataaa gccgcccaag 720
tcttcgcctt gctttttgtt acagagtatc tcactaagtg gcccaagttt ttttttgaca 780
ttctctcagt agtggaccta aatccaaggg gagtagatct ctacctgcga atcctcatgg 840
ctattgattc agagttggtg gatcgtgatg tggtgcatac atcagaggag gctcgtagga 900

atactctcat aaaagatacc atgagggaac agtgcattcc aaatctggtg gaatcatggt 960
accaaatatt acaaaattat cagtttacta attctgaagt gacgtgtcag tgccttgaag 1020
tagttggggc ttatgtctct tggatagact tatcccttat agccaatgat aggtttataa 1080
atatgctgct aggtcatatg tcaatagaag ttctacggga agaagcatgt gactgtttat 1140
ttgaagttgt aaataaagga atggaccctg ttgataaaat gaaactagtg gaatctttgt 1200
gtcaagtatt acagtctgct gggtttttca gcattgacca ggaagaagat gttgacttcc 1260
tggccagatt ttctaagttg gtaaatggaa tgggacagtc attgatagtt agttggagta 1320
aattaattaa gaatggggat attaagaatg ctcaagaggc actacaagct attgaaacaa 1380
aagtggcact gatgttgcag ctactaattc atgaggatga tgatatttct tctaatatta 1440
ttggattttg ttacgattat cttcatattt tgaaacagct tacagtgctc tcggatcagc 1500
aaaaagctaa tgtagaggca atcatgttgg ccgttatgaa aaaattgact tacgatgaag 1560
aatataactt tgaaaatgag ggtgaagatg aagccatgtt tgtagaatat agaaaacaac 1620
tgaagttact gttggacagg cttgctcaag tttcaccaga gttactactg gcctctgttc 1680
gcagagtttt tagttctaca ctgcagaatt ggcagactac acggtttatg gaagttgaag 1740
tagcaataag attgctgtat atgttggcag aagctcttcc agtatctcat ggtgctcact 1800
tctcaggtga tgtttcaaaa gctagtgctt tgcaggatat gatgcgaact ctggtaacat 1860
caggagtcag ttcctatcag catacatctg tgacattgga gttcttcgaa actgttgtta 1920
gatatgaaaa gtttttcaca gttgaacctc agcacattcc atgtgtacta atggctttct 1980
tagatcacag aggtctgcgg cattccagtg caaaagttcg gagcaggacg gcttacctgt 2040
tttctagatt tgtcaaatct ctcaataagc aaatgaatcc tttcattgag gatattttga 2100
atagaataca agatttatta gagctttctc cacctgagaa tggccaccag tccttactga 2160
gcagcgatga tcaacttttt atttatgaga cagctggagt gctgattgtt aatagtgaat 2220
atccggcaga aaggaaacaa gccttaatga ggaatctgtt gactccacta atggagaagt 2280
ttaaaattct gttagaaaag ttgatgctgg cacaagatga agaaaggcaa gcctctctag 2340
cagactgtct taaccatgct gttggatttg caagtcgaac cagtaaagct ttcagcaaca 2400
aacagactgt gaaacaatgt ggctgttccg aagtttatct ggactgttta cagacattct 2460
tgccagccct cagttgtccc ttacaaaagg atattctcag aagtggagtc cgtactttcc 2520
ttcatcgaat gattatttgc ctggaggaag aagttcttcc gttcattcca tctgcttcag 2580
aacatatgct caaagattgt gaagcaaaag atctccagga gttcattcct cttatcaacc 2640
agattacggc caaattcaag atacaggtat ccccgttttt acaacagatg ttcatgcccc 2700
tgcttcatgc aatttttgaa gtgctgctcc ggccagcaga agaaaatgac cagtctgctg 2760
ctttagagaa gcagatgttg cggaggagtt actttgcttt cctgcaaaca gtcacaggca 2820
gtgggatgag cgaagttata gcaaatcaag gtgcagagaa tgtagaaaga gtgttggtta 2880
ctgttatcca aggagcagtt gaatatccag atccaattgc acagaaaaca tgttttatca 2940
tcctctcaaa gttggtagaa ctctggggag gtaaagatgg accagtggga tttgctgatt 3000
ttgtttataa gcacattgtc cccgcatgtt tcctagcacc tttaaaacaa acctttgacc 3060
tggcagatgc acaaacagta ttggctttat ctgagtgtgc agtgacactg aaaacaattc 3120
atctcaaacg gggcccagaa tgtgttcagt atcttcaaca agaatacctg ccctccttgc 3180
aagtagctcc agaaataatt caggagtttt gtcaagcgct tcagcagcct gatgctaaag 3240
tttttaaaaa ttacttaaag gtgttcttcc agagagcaaa gccctgagga ctggatttcc 3300
ctgtgcctac ttcatgatca tgaattccag ttaatttata aagaggcgat ttttgtgtgc 3360
cattcacact ggtctttttc acattgtttt gagcttattg cagtatatgt tttgggattt 3420
ttctgtaaaa tgggtgtaat tttcctaata caggtatgta acaacaaaag aagttgcctg 3480
catgccggtc caaattgttc tgtataaaga tgctcttaaa agacacaaga gttatcctag 3540
aaccttaatt ctttttttatt tgaaatttta agtcaagtcc tttataaaga ccatagcagt 3600
ggaaaacagt gtacttttta aaaaattgct gaatataaaa tctttgaaaa ttttctttat 3660
gtgtgaagac acaaagtatg ggggaagaca gcaatcaaaa ctaacttttt gtagatagcc 3720
atttcatttc tttaaactgt ttcaacgcca atatgtattc tacaaaagag aatggtttta 3780
ggctccagtg ttatactttt ttttatatat atatataaaa ataaacttta cgtt 3834

<210> 28
<211> 2178
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte Identification No.: 2263514CB1

<400> 28
cagctcccta agcggttgtc accgctggag acggttggga gaaccgttgt ggcgagcgct 60
acacgaggca aacgacttct cccttctttg aactggaccc cgcgagcacc agagtcggcg 120
taactatcgc ctgacaggca tttaaatcaa acggtattga gatggattgg gttatgaaac 180
ataatggtcc aaatgacgct agtgatggga cagtacgact tcgtggacta ccatttggtt 240
gcagcaaaga ggaaatagtt cgagttcttt caaggtatat tgagatcttc agaagtagca 300
ggagtgaaat caaaggattt tatgatccac caagaagatt gctgggacag cgaccgggac 360
catatgatag accaatagga ggaagagggg gttattatgg agctgggcgt ggaagttatg 420
gaggttttga tgactatggt ggctataata attacggcta tgggaatgat ggctttgatg 480
acagaatgag agatggaaga ggtatgggag gacatggcta tggtggagct ggtgatgcaa 540
gttcaggttt tcatggtggt catttcgtac atatgagagg gttgcctttt cgtgcaactg 600
aaaatgccat tgctaatttc ttctcaccac taaatccaat acgagttcat attgatattg 660
gagctgatgg cagagccaca ggagaagcag atgtagagtt tgtgacacat gaagatgcag 720
tagctgccat gtctaaagat aaaaataaca tgcaacatcg atatattgaa ctcttcttga 780
attctactcc tggaggcggc tctggcatgg gaggttctgg aatgggaggc tacggaagag 840
atggaatgga taatcaggga ggctatggat cagttggaag aatgggaatg gggaacaatt 900
acagtggagg atatggtact cctgatggtt tgggtggtta tggccgtggt ggtggaggca 960
gtggaggtta ctatgggcaa ggcggcatga gtggaggtgg atggcgtggg atgtactgaa 1020
agcaaaaaca ccaacataca agtcttgaca acagcatctg gtctactaga ctttcttaca 1080
gatttaattt cttttgtatt ttaagaactt tataatgact gaaggaatgt gttttcaaaa 1140
tattatttgg taaagcaaca gattgtgatg ggaaaatgtt ttctgtaggt ttatttgttg 1200
catactttga cttaaaaata aatttttata ttcaaaccac tgatgttgat actttttata 1260
tactagttac tcctaaagat gtgctgcctt cataagattt gggttgatgt attttactat 1320
tagttctaca agaagtagtg tggtgtaatt ttagaggata atggttcacc tctgcgtaaa 1380
ctgcaagtct taagcagaca tctggaatag agcttgacaa ataattagtg taacttttt 1440
ctttagttcc tcctggacaa cactgtaaat ataaagccta aagatgaagt ggcttcagga 1500
gtataaattc agctaattat ttctatatta ttatttttca aatgtcattt atcaggcata 1560
gctctgaaac attgatgatc taagaggtat tgatttctga atattcataa ttgtgttacc 1620
tgggtatgag agtgttggaa gctgaattct agccctagat tttggagtaa aacccttca 1680
gcacttgacc gaaataccaa aaatgtctcc aaaaaattga tagttgcagg ttatcgcaag 1740
atgtcttaga gtagggttaa ggttctcagt gacacaagaa ttcagtatta agtacatagg 1800
tatttactat ggagtataat tctcacaatt gtattttcag ttttctgccc aatagagttt 1860
aaataactgt ataaatgatg actttaaaaa aatgtaagca acaagtccat gtcatagtca 1920
ataaaaacaa tcctgcagtt gggttttgta tctgatccct gcttggagtt ttagtttaaa 1980
gaatctatat gtagcaagga aaaggtgctt tttaatttta atccctttga tcaatatggc 2040
ttttttccaa attggctaat ggatcaaaat gaaacctgtt gatgtgaatt cagttattga 2100
acttgttact tgtttttgcc agaaatgtta ttaataaatg tcaatgtggg agataaaaaa 2160
aaaaaaaaaa aaactggg 2178

<210> 29
<211> 1503
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2738270CB1

<400> 29
acgacctttt ggccaggtta gggaggggge gacgctgaga tgggggcggc ggcggcggaa 60
gcggatcgca ctctctttgt gggcaacctt gaaacgaaag tgaccgagga gctccttttc 120
gagcttttcc accaggctgg gccagtaata aaggtgaaaa ttccaaaaga taaggatggt 180
aaaccaaagc agtttgcgtt tgtgaatttc aaacatgaag tgtctgttcc ttatgcaatg 240

aatctactta atggaatcaa actttatgga aggcctatca aaattcaatt tagatcagga 300
agtagtcatg ccccacaaga tgtcagtttg tcatatcccc aacatcatgt tggaaattca 360
agccctacct ccacatctcc tagcagcagg tacgaaagga ctatggataa catgacttca 420
tcagcacaga taattcagag atctttctct tctccagaaa attttcagag acaagcagtg 480
atgaacagtg ctttgagaca aatgtcatat ggtggaaaat ttggttcttc acctctggat 540
caatcaggat tttcaccatc agttcaatca cacagtcata gtttcaatca gtcttcaagc 600
tcccagtggc gccaaggtac accatcatca cagcgtaaag tcagaatgaa ttcttatccc 660
tacctagcag atagacatta tagccgggaa cagcgttaca ctgatcatgg gtctgaccat 720
cattacagag gaaagagaga tgatttcttc tatgaagaca ggaatcatga tgactggagc 780
catgactatg ataacagaag agacagtagt agagatggaa aatggcgctc atctcgacac 840
taacacatgt taaaaggaca ttgtttttat agggtcattt taggcccttt gactaagttg 900
atatggaaat attttgttga aaaactgtac agagcagctt tacaagttgt cacattttct 960
ttataaattt ttttaaagct acagtttaat acaaaatgaa ttgcggtttt attacattaa 1020
taacctttca cctcagggtt ttatgaagag gaaagggttt tatgcaaaag aaagtgctac 1080
aattcctaat cattttagac actttaggag ggggtgaagt tgtatgataa agcagatatt 1140
ttaattattt gttatctttt tgtattgcaa gaaatttctt gctagtgaat caagaaaaca 1200
tccagattga cagtctaaaa tggctactgg tattttagtt aattcaaaaa tgaaactttt 1260
cagtgattca ctttactaac attctatttg agaaggctta ttggtaaagt ttggggataa 1320
aggcattgct taacttctta tataatttag gtataaattc tgtgacatgc tcttgagctt 1380
taccctagtt gaacatacat gtgtagattt acacatactg tttcattcta aaaattttag 1440
aattgttcat taaaacccca tttgaggtat aaggtcactc aggaaggtta aaatatctcc 1500
acc 1503

<210> 30
<211> 2548
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2824412CB1

<400> 30
tctgggagtc cgaccaggca tgcctcctca gcctcagggg cctgcaccct tacgtcgtcc 60
tgactcatct gatgaccgtt atgtaatgac aaaacatgcc accatttatc caactgaaga 120
ggagttacag gcagttcaga aaattgtttc tattactgaa cgtgctttaa aactcgtttc 180
agacagtttg tctgaacatg agaagaacaa gaacaaagag ggagatgata agaaagaggg 240
aggtaaagac agagctttga aaggagtttt gcgagtggga gtattggcaa aaggattact 300
tctccgagga gatagaaatg tcaaccttgt tttgctgtgc tcagagaaac cttcaaagac 360
attattaagc cgtattgcag aaaacctacc caaacagctt gctgttataa gccctgagaa 420
gtatgacata aaatgtgctg tatctgaagc ggcaataatt ttgaattcat gtgtggaacc 480
caaaatgcaa gtcactatca cactgacatc tccaattatt cgagaagaga acatgaggga 540
aggagatgta acctcgggta tggtgaaaga cccaccggac gtcttggaca ggcaaaaatg 600
ccttgacgct ctggctgctc tacgccacgc taagtggttc caggctagag ctaatggtct 660
gcagtcctgt gtgattatca tacgcattct tcgagacctc tgtcagcgag ttccaacttg 720
gtctgatttt ccaagctggg ctatggagtt actagtagag aaagcaatca gcagtgcttc 780
tagccctcag agccctgggg atgcactgag aagagttttt gaatgcattt cttcagggat 840
tattcttaaa ggtagtcctg gacttctgga tccttgtgaa aaggatccct ttgatacctt 900
ggcaacaatg actgaccagc agcgtgaaga catcacatcc agtgcacagt ttgcattgag 960
actccttgca ttccgccaga tacacaaagt tctaggcatg gatccattac cgcaaatgag 1020
ccaacgtttt aacatccaca acaacaggaa acgaagaaga gatagtgatg gagttgatgg 1080
atttgaagct gaggggaaaa aagacaaaaa agattatgat aacttttaaa aagtgtctgt 1140
aaatcttcag tgttaaaaaa acagatgccc atttgttggc tgtttttcat tcataataat 1200
gtctacattg aaaaatttat caagaattta aaggatttca tggaagaacc aagtttttct 1260
atgatattaa aaaatgtaca gtgttaggta ttatttgaat ggaaagacac ccaaaaaaaa 1320
aaatgtgctc cgactagggg gaaaacagta gttccgattt tttcccatta tttttatttt 1380

attttctggt tgccctagct tccccccccta tttttgtgtc ttttattaac tagtgcattg 1440
tcttattaaa tcttcactgt atttaatgca ggatgtgtgc ttcagttgct ctgtgtattt 1500
tgatatttta atttagaggt tttgtttgct ttttgacact agttgtaagt tactttgtta 1560
tagatggtat cctttacccc ttcttaatat tttacagcag tacgtttttt tgtaacgtga 1620
gactgcagag tttgtttttc tatatgtgaa ggattacaac acaaaaagtt atcctgccat 1680
tcgagtgctc agaactgaat gtttctgcag atcttgtggc atttgtctct agtgtgatat 1740
ataaaggtgt aattaagaca gagttctgtt aatctaatca agtttgctgt tagttgtgca 1800
ttagcagtat aaaagctaat atatactata tggtcttgca acagttttaa agcctctgca 1860
taattgataa taaaaatgca tgacattctt gtttttaata gacttttaaa atcataattt 1920
taggtttaac acgtagatct ttgtacagtt gacttttttga catagcaagg ccaaaaataa 1980
ctttctgaat attttttttct tgtgtataag tggaaagggc atttttcaca tataagtggg 2040
ctaaccaata ttttcaaaag aacttcatca ttgtacaact aacaacagta actagccctt 2100
aattatggtg acagttcctt attggtgtgt gtgagattac tctagcaact attacagtat 2160
aacacagatg atcttctcca cacaccccat cacccagata atttacagtt ctgttaacag 2220
tgaggttgat aaagtattac tgataaaaaa ttatctaagg aaaaaaacag aaaattattt 2280
ggtgtggcca tcttacctgc ttatgtctcc tacacaaagc taaatattct agcagtgatg 2340
taatgaaaaa ttacatctta ctgttgatat atgtatgctc tggtacacag atgtcatttt 2400
gttgtcacag cactacagtg aaatacacaa aaaatgaaat tcatataatg acttaaatgt 2460
attatatgtt agaattgaca acataaacta cttttgcttt gaaatgatgt atgcttcagt 2520
aaaatcatat tcaaatttaa aaaaaaaa 2548

<210> 31
<211> 811
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 002690CB1

<400> 31
cgcatgcgtt ccctgaaatt gccgccaccg gctctacctt ccagtttcca gttccggcct 60
ccaaggggcg ggcagaagtt ggaaacatgc ggctgtcggt cgctgcagcg atctcccatg 120
gccgcgtatt tcgccgtatg ggcctcggtc ccgagtcccg catccatctg ttgcggaact 180
tgctcacagg gctggtgcgg cacgaacgca tcgaggcacc atgggcgcgt gtggacgaaa 240
tgaggggcta cgcggagaag ctcatcgact atgggaagct gggagacact aacgaacgag 300
ccatgcgcat ggctgacttc tggctcacag agaaggattt gatcccaaag ctgtttcaag 360
tactggcccc tcggtacaaa gatcaaactg ggggctacac aagaatgctg cagatcccaa 420
atcggagttt ggatcgggcc aagatggcag tgatcgagta taaagggaat tgcctcccac 480
ccctgcctct gcctcgcaga gacagccacc ttacactcct aaaccagctg ctgcagggtt 540
tgcggcagga cctcaggcaa agccaggaag caagcaacca cagctcccac acagctcaaa 600
caccagggat ttaactggat ctgaagagtc tgcagccctt aatcagtacc catgatcaca 660
ggcctttgga gcacttttac tctctgagaa gaactggagc tagagatgta aaatggacag 720
tcttgatggg gttgagaacc ttctggggag ccagatgacc ctctctttgc acaatagata 780
aaagtcttta tatgaatata aaaaaaaaaa a 811

<210> 32
<211> 1457
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 041108CB1

<400> 32
tacggaagct gggtcttctt gctgtgaggt cgcgttcccc agtgttacgg agggtccttg 60
aggcaggagt gaaaattggg tctgggggtt agtcctgggg tggaggtctg ggcacgccgg 120
gtcggacccc ctccatcttc ggttttgcac accccgcttt ccagcgcgga gtcgcggcgg 180
gtagggcggc gtcgcgtgcg tgacgtcatc cagcggcgcc tcgcgaggct ccagtggcct 240
tgacctcccg cggcgtggga ggctgcgcgg cgatgctgca gttcgtccgg gccggggcgc 300
gggcctggct tcggcctacc ggcagccagg gcctgagttc cctggcggaa gaggcagcgc 360
gtgcgaccga gaacccggag caggtggcga gcgagggtct cccggagccc gtgctgcgca 420
aagtcgagct cccggtaccc actcatcgac gcccagtgca ggcctgggtc gagtccttgc 480
ggggcttcga gcaggagcgc gtgggcctgg ccgacctgca ccccgatgtt ttcgccaccg 540
cgcccaggct ggacatactg caccaggttg ctatgtggca gaagaacttc aagagaatta 600
gctatgccaa gaccaagacg agagccgagg tgcggggcgg tggccggaag ccttggccgc 660
agaaaggcac tgggcgggcc cggcatggca gcatccgctc tccgctctgg cgaggaggag 720
gtgttgccca tggcccccgg ggccccacaa gttactacta catgctgccc atgaaggtgc 780
gggcgctggg tctcaaagtg gcactgaccg tcaagctggc ccaggacgac ctgcacatca 840
tggactccct agagctgccc accggagacc cacagtacct gacagagctg gcgcactacc 900
gccgctgggg ggactccgta ctcctcgtgg acttaacaca cgaggagatg ccacagagca 960
tcgtggaggc cacctctagg cttaagacct tcaacttgat cccggctgtt ggcctaaatg 1020
tgcacagcat gctcaagcac cagacgctgg tcctgacgct gcccaccgtc gccttcctgg 1080
aggacaagct gctctggcag gactcacgtt acagaccccc ctacccccttc agcctgccct 1140
acagcgactt cccccgaccc ctaccccacg ctacccaggg cccagcggcc accccgtacc 1200
actgttgatg tgaagcacct cttctgagcc aggccgagcc cctggccgac ttgggagcct 1260
caggcccacg cccacccttc gaggaaggtg tcacctggac cccttcattc cacggaggaa 1320
gctgaggcca cagggagcgg ccatcgccat tgggaagggg cgactccacg gagagcccag 1380
acgggcttct gcatccattc cctcttttttg tttttaaaat aaattgtatt tttgaatcaa 1440
ggaggaaaaa aaaaaaa 1457

<210> 33
<211> 1357
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 869138CB1

<400> 33
ctggtggcgt tcaagatgtc gaccaagaat ttccgagtca gtgacgggga ctggatttgc 60
cctgacaaaa aatgtggaaa tgtaaacttt gctagaagaa ccagctgtaa tcgatgtggt 120
cgggagaaaa caactgaggc caagatgatg aaagctgggg gcactgaaat aggaaagaca 180
cttgcagaaa agagccgagg cctatttagt gctaatgact ggcaatgtaa aacttgcagc 240
aatgtgaatt gggccagaag atcagagtgt aatatgtgta atactccaaa gtatgctaaa 300
ttagaagaaa gaacaggata tggtggtggt tttaatgaaa gagaaaatgt tgaatatata 360
gaaagagaag aatctgatgg tgaatatgat gagtttggac gtaaaaagaa aaaatacaga 420
gggaaagcag ttggtcctgc atctatatta aaggaagttg aagataaaga atcagaggga 480
gaagaagagg atgaggatga agatctttct aaatataagt tagatgagga tgaggatgaa 540
gatgacgctg atctctcaaa atataatctt gatgccagtg aagaagaaga tagtaataaa 600
aagaaatcta atagacgaag tcgctcaaag tctcgatctt cacattcacg atcttcatca 660
cgctcatcct ccccctcaag ttcaaggtct aggtccaggt cccgttcaag aagttcttcc 720
agttcgcagt caagatctcg ttccagttcc agagaacgtt cgagatctcg tgggtcgaaa 780
tcaagatcca gctccaggtc ccacaggggc tcttcttccc cacgaaaaag atcttattca 840
agttcatcat cttctcctga gaggaacaga aagagaagtc gttctagatc ttcttcatct 900
ggtgatcgca aaaaaagacg aacaagatca cggtcacccg aaagacgcca caggtcatca 960
tctggatcat cccattctgg ttcccgttca agttcaaaaa agaaataatg tattaaaatt 1020
tacatcttaa aaaaatccag tacagtgcat gaagcatatt tttaaagaag ttggtgtctt 1080
acttggtcag aagtgctaaa tctgctagta gaggtgcatg cctttcattg cttttcaaaa 1140

caatacagct gtgtttattt gtgaagttaa aagtaaatag cattttaagc cataatgtcc 1200
caaaatagat gttctgtcat tcattattta caaccatttg cttcatttaa aaccatttca 1260
gctataacaa agtactttgc ttcctaattt aaacccattt ttgtcatttc caaatacatc 1320
ctgtccattg gctaagacag gattacctag gcttgct 1357

<210> 34
<211> 1326
<212> DNA
<213> Homo sapiens

<221> unsure
<222> 1313
<223> a or t or g or c, unknown, or other

<220>
<221> misc_feature
<223> Incyte Identification No.: 934406CB1

<400> 34
ttttcttcgg ggactatcct tgtctgatca ggcgggaaag acggtgccgc ccgacaatgc 60
gcggaggtag gaggggggaag tggaggcggg agtgaagtct cgcgagaaga gtcggttgcc 120
gtagcagagc cctctagctg tgtgtgtctg aggctcggcc gcctgagccg cggacggttt 180
gctgagcccg ttagtgcgcc cggccgagac acgccgccgc catgtcccgc tacctgcgtc 240
cccccaacac gtctctgttc gtcaggaacg tggccgacga caccaggtct gaagacttgc 300
ggcgtgaatt tggtcgttat ggtcctatag ttgatgtgta tgttccactt gatttctaca 360
ctcgccgtcc aagaggattt gcttatgttc aatttgagga tgttcgtgat gctgaagacg 420
ctttacataa tttggacaga aagtggattt gtggacggca gattgaaata cagtttgccc 480
agggggatcg aaagacacca aatcagatga aagccaagga agggaggaat gtgtacagtt 540
cttcacgcta tgatgattat gacagataca gacgttctag aagccgaagt tatgaaagga 600
ggagatcaag aagtcggtct tttgattaca actatagaag atcgtatagt cctagaaaca 660
gtagaccgac tggaagacca cggcgtagca gaagccattc cgacaatgat agaccaaact 720
gcagctggaa tacccagtac agttctgctt actacacttc aagaaagatc tgaaagcgga 780
aaaagaacca aagaagggca gttcaagcga ccaaagggtg ggtggaaggt gctgcagtat 840
gaatactgta cgaatatttt gactctggtc tgaaaagata aaagaatgtt atcgaaaact 900
acatggaata attgaagtcc cttcaagttt gaaagtaagc attttaggac aaataaaagg 960
aaattcaact ttgtacttgt ggaaactaat ccctaaatat gaataggttt atattgattc 1020
atgggtaaca ggtccataat aaattattgg aaactaggat gtctgaatat caaggaagac 1080
agccatagtc tcttacagtg cctctgttgg tctgtctcaa actgaattgg gtgggaaaag 1140
gtatggtcca atataaaagt tccatttttg ccattattgg gcaaatcttg cctttgttta 1200
ttttggtgcc agtgttttct gcttaatcat ttgctttgtt ggcatctgtg tttatttact 1260
tgtacaccac atgcagttta catctgtctt aactactcct tcccaggtaa ttnccattat 1320
attgac 1326

<210> 35
<211> 3301
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1315083CB1

<400> 35
gaagggaagt aacgtcagcc tgagaactga gtagctgtac tgtgtggcgc cttattctag 60
gcacttgttg ggcagaatgt cacacctgcc gatgaaactc ctgcgtaaga agatcgagaa 120

gcggaacctc aaattgcggc agcggaacct aaagtttcag ggggcctcaa atctgaccct 180
atcggaaact caaaatggag atgtatctga agaaacaatg ggaagtagaa aggttaaaaa 240
atcaaaacaa aagcccatga atgtgggctt atcagaaact caaaatggag gcatgtctca 300
agaagcagtg ggaaatataa aagttacaaa gtctccccag aaatccactg tattaaccaa 360
tggagaagca gcaatgcagt cttccaattc agaatcaaaa aagaaaaaga agaaaaagag 420
aaaaatggtg aatgatgctg agcctgatac gaaaaaagca aaaactgaaa acaaagggaa 480
atctgaagaa gaaagtgccg agactactaa agaaacagaa aataatgtgg agaagccaga 540
taatgatgaa gatgagagtg aggtgcccag tctgcccctg ggactgacag gagcttttga 600
ggatacttcg tttgcttctc tatgtaatct tgtcaatgaa aacactctga aggcaataaa 660
agaaatgggt tttacaaaca tgactgaaat tcagcataaa agtatcagac cacttctgga 720
aggcagggat cttctagcag ctgcaaaaac aggcagtggt aaaaccctgg cttttctcat 780
ccctgcagtt gaactcattg ttaagttaag gttcatgccc aggaatggaa caggagtcct 840
tattctctca cctactagag aactagccat gcaaaccttt ggtgttctta aggagctgat 900
gactcaccac gtgcatacct atggcttgat aatgggtggc agtaacagat ctgctgaagc 960
acagaaactt ggtaatggga tcaacatcat tgtggccaca ccaggccgtc tgctggacca 1020
tatgcagaat accccaggat ttatgtataa aaacctgcag tgtctggtta ttgatgaagc 1080
tgatcgtatc ttggatgtgg ggtttgaaga ggaattaaag caaattatta aacttttgcc 1140
aacacgtaga cagactatgc tcttttctgc cacccaaact cgaaaagttg aagacctggc 1200
aaggatttct ctgaaaaagg agccattgta tgttggcgtt gatgatgata aagcgaatgc 1260
aacagtggat ggtcttgaac agggatatgt tgtttgtcct tctgaaaaga gattccttct 1320
gctctttaca ttccttaaga agaaccgaaa gaagaagctt atggtcttct tttcatcttg 1380
tatgtctgtg aaataccact atgagttgct gaactacatt gatttgcccg tcttggccat 1440
tcatggaaag caaaagcaaa ataagcgtac aaccacattc ttccagttct gcaatgcaga 1500
ttcgggaaca ctattgtgta cggatgtggc agcgagagga ctagacattc ctgaagtcga 1560
ctggattgtt cagtatgacc ctccggatga ccctaaggaa tatattcatc gtgtgggtag 1620
aacagccaga ggcctaaatg ggagagggca tgccttgctc attttgcgcc cagaagaatt 1680
gggttttctt cgttacttga aacaatccaa ggttccatta agtgaatttg acttttcctg 1740
gtctaaaatt tctgacattc agtctcagct tgagaaattg attgaaaaga attactttct 1800
tcataagtca gcccaggaag catataagtc atacatacga gcctatgatt cccattctct 1860
gaaacagatc tttaatgtta ataacctaaa tttgcctcag gttgctctgt catttggttt 1920
caaggtgcct cccttcgttg atctgaacgt caacagtaat gaaggcaagc agaaaaagcg 1980
aggaggtggt ggtggatttg gctaccagaa aaccaagaaa gttgagaaat ccaaaatctt 2040
taaacacatt agcaagaaat catctgacag caggcagttc tctcactgaa cacatgcctt 2100
cctttcatct tgaataactt tgtcctaaaa tgaatttttt ttccccttga tttaacagga 2160
tttttgtaga ctttagaatt tggacttacc taacaagagt ataaattgac ttgggttgca 2220
agcactgagc actgttactt ctatcacgtc tctcttttat ttctgggata taaaacaggc 2280
tttaagtttc ttggttgccc aagggcagag caaggaatat ctggtgtttc ttgtgatgat 2340
aatattttaa ttttaaatat ccctccctca tacaagtgta tgttaccatt ttaatataat 2400
tctttttgta cctttccttc ttgttttgtg aagatttttg tggcatggat tgctgtgctc 2460
actgctgtaa aaggtgacct agtgtactgg gcagctggtg gcggtgcaga aaagagtctc 2520
aggttatttt ttgtttttag ttatttcttg gaccttgaca gtatctaatg actcctcctg 2580
aaaatgctgc agtataaaag agcaaagagc tttgggaaat acctaagaag caccttaaga 2640
ttagggtggc attgctttta tagattcttg attttaaagc aacaggcctt tctcaggtgt 2700
tgcatttttt ggagcaaaaa ctatgggttg taatttgaat aaagtgtcac taagcagtta 2760
taacgtttga tggctggggg gtaggaagag gatggaattg agatgtttga gcctcattta 2820
catcaataga ggtgtaatgt actgcatttc ttcatttggt aacataacaa agactttcat 2880
acaaagaacg atgatgctcc tcattaagat ttgtttaatt caaggtggtt tggatttggt 2940
aagcctttgc actctgtaga gtacttagaa gacaagggca acttacttgg agttagagcc 3000
aagctgtcag acggtgccca gcacacatta atgttagctt ctttctgaga aaaaaatacc 3060
tcttccaggc cctgaaacaa aaaatacatt tgctgtgaag attgaaaatg aacaaagtta 3120
gaaaaaaaaa cagcaaaatc agtgatttag tcagatgagt ttttcgttgt aggagcactt 3180
gatttctagt gtgttttgta cagtatataa ctacaagata gtacattttg tagcagttca 3240
aagccaaagt tgctagcatc attttgctgt tgtgccagtt aatcatagga tcccattaag 3300
g 3301

<210> 36

<211> 1703
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1444908CB1

<400> 36
gcttcccctc cgtcatccat cgctcggcgg tcttctcttc tccatgggtc tgctctgcgc 60
gttccatcga gtttctcggc catcgcgcgc ctgcgccatt gggctgtcag tcagaggcgg 120
cgtggagatc gctgggagcg gttgcggcgt gcggggagct gagttatagc tgtgacttct 180
gccctgccag gccgcacaca agctggctga cccggtttgt aaaaatggaa tttcaagcag 240
tagtgatggc agtaggtgga ggatctcgga tgacagacct aacttccagc attcccaaac 300
ctctgcttcc agttgggaac aaacctttaa tttggtaccc attgaacctg cttgagcgtg 360
ttggatttga agaagtcatt gtggttacaa ccagggatgt tcaaaaggct ctatgtgcag 420
aattcaagat gaaaatgaag ccagatattg tgtgtattcc tgatgatgct gacatgggaa 480
ctgcagattc tttgcgctac atatatccaa aacttaagac agatgtgctg gtgctgagct 540
gtgatctgat aacagacgtt gccttacatg aggttgtgga cctgtttaga gcttatgatg 600
catcacttgc tatgttgatg agaaaaggcc aagatagcat agaacctgtt cccggtcaaa 660
aggggaaaaa aaaagcagtg gagcagcgtg acttcattgg agtggacagc acaggaaaga 720
ggctgctctt catggctaat gaagcagact tggatgaaga gctggtcatt aagggatcca 780
tcctacagaa gcatcctaga atacgtttcc acacgggtct tgtggatgcc cacctctact 840
gtttgaaaaa atacatcgtg gatttcctaa tggaaaatgg gtcaataact tctatccgga 900
gtgaactgat tccatattta gtgagaaaac agttttcctc agcttcctca caacagggac 960
aagaagaaaa agaggaggat ctaaagaaaa aggagctgaa gtccttagat atctacagtt 1020
ttataaaaga agccaataca ctgaacctgg ctccctatga tgcctgctgg aatgcctgtc 1080
gaggagacag gtgggaagac ttgtccagat cacaggtgcg ctgctatgtc cacatcatga 1140
aagaggggct ctgctctcga gtgagcacac tgggactcta catggaagca aacagacagg 1200
tgcccaaatt gctgtctgct ctctgtccag aagaaccacc agtccattcg tcagcccaga 1260
ttgtcagcaa acacctggtt ggagttgaca gcctcattgg gccagagaca cagattggag 1320
agaagtcatc cattaagcgc tcagtcattg gctcatcctg tctcataaaa gatagagtga 1380
ctattaccaa ttgccttctc atgaactcag tcactgtgga ggaaggaagc aatatccaag 1440
gcagtgtcat ctgcaacaat gctgtgatcg agaagggtgc agacatcaag gactgcttga 1500
ttggaagtgg ccagaggatt gaagccaaag ctaaacgagt gaatgaggtg atcgtgggga 1560
atgaccagct catggagatc tgagttctga gcaagtcaga ctccttcctt ttggcctcca 1620
aagccacaga tgttggccgg cccacctgtt taactctgta tttatttccc aataaagaag 1680
ggcttccaaa ggcaaaaaaa aaa 1703

<210> 37
<211> 2536
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1557481CB1

<400> 37
gacttccggg gcggcggttg catcagattc taggaagtgt ctgtagccgc agctgcgggt 60
ccgggattcc cagccatggc agattcctcc gggcagcagg gcaaaggccg gcgtgtgcag 120
ccccagtggt cccctcctgc tgggacccag ccatgcagac tccaccttta caacagcctc 180
accaggaaca aggaagtgtt catacctcaa gatgggaaaa aggtgacgtg gtattgctgt 240
gggccaaccg tctatgacgc atctcacatg gggcacgcca ggtcctacat ctcttttgat 300
atcttgagaa gagtgttgaa ggattacttc aaatttgatg tcttttattg catgaacatt 360
acggatattg atgacaagat catcaagagg gcccggcaga accacctgtt cgagcagtat 420

cgggagaaga ggcctgaagc ggcacagctc ttggaggatg ttcaggccgc cctgaagcca 480
ttttcagtaa aattaaatga gaccacggat cccgataaaa agcagatgct cgaacggatt 540
cagcacgcag tgcagcttgc cacagagcca cttgagaaag ctgtgcagtc cagactcacg 600
ggagaggaag tcaacagctg tgtggaggtg ttgctggaag aagccaagga tttgctctct 660
gactggctgg attctacact tggctgtgat gtcactgaca attccatctt ctccaagctg 720
cccaagttct gggaggggga cttccacaga gacatggaag ctctgaatgt tctccctcca 780
gatgtcttaa cccgggttag tgagtatgtg ccagaaattg tgaactttgt ccagaagatt 840
gtggacaacg gttacggcta tgtctccaat gggtctgtct actttgatac agcgaagttt 900
gcttctagcg agaagcactc ctatgggaag ctggtgcctg aggccgttgg agatcagaaa 960
gcccttcaag aaggggaagg tgacctgagc atctctgcag accgcctgag tgagaagcgc 1020
tctcccaacg actttgcctt atggaaggcc tctaagcccg gagaaccgtc ctggccgtgc 1080
ccttggggaa agggtcgtcc gggctggcat atcgagtgct cggccatggc aggcaccctc 1140
ctaggggctt cgatggacat tcacggaggt gggttcgacc tccggttccc ccaccatgac 1200
aatgagctgg cacagtcgga ggcctacttt gaaaacgact gctgggtcag gtacttcctg 1260
cacacaggcc acctgaccat tgcaggctgc aaaatgtcaa agtcactaaa aaacttcatc 1320
accattaaag atgccttgaa aaagcactca gcacggcagt tgcggctggc cttcctcatg 1380
cactcgtgga aggacaccct ggactactcc agcaacacca tggagtcagc gcttcaatat 1440
gagaagttct tgaatgagtt tttcttaaat gtgaaagata tccttcgcgc tcctgttgac 1500
atcactggtc agtttgagaa gtggggagaa gaagaagcag aactgaataa gaacttttat 1560
gacaagaaga cagcaattca caaagccctc tgtgacaatg ttgacacccg caccgtcatg 1620
gaagagatgc gggccttggt cagtcagtgc aacctctata tggcagcccg gaaagccgtg 1680
aggaagaggc ccaaccaggc tctgctggag aacatcgccc tgtacctcac ccatatgctg 1740
aagatctttg gggccgtaga agaggacagc tccctgggat tcccggtcgg agggcctgga 1800
accagcctca gtctcgaggc cacagtcatg ccctaccttc aggtgttatc agaattccga 1860
gaaggagtgc ggaagattgc ccgagagcaa aaagtccctg agattctgca gctcagcgat 1920
gccctgcggg acaacatcct gcccgagctt ggggtgcggt ttgaagacca cgaaggactg 1980
cccacagtgg tgaaactggt agacagaaac accttattaa aagagagaga agaaaagaga 2040
cgggttgaag aggagaagag gaagaagaaa gaggaggcgg cccggaggaa acaggaacaa 2100
gaagcagcaa agctggccaa gatgaagatt ccccccagtg agatgttctt gtcagaaacc 2160
gacaaatact ccaagtttga tgaaaatggt ctgcccacac atgacatgga gggcaaagag 2220
ctcagcaaag ggcaagccaa gaagctgaag aagctcttcg aggctcagga gaagctctac 2280
aaggaatatc tgcagatggc ccagaatgga agcttccagt gaggggggcac aggactgact 2340
ttttaaacca ttgtggacta gtggctgctg tctgcctcag tgacaatgtc ccagcgctcc 2400
tatcatgttt acagtcaccc ttgggtccta aattaagagt tgtgttcatg taggttcgtg 2460
tcgtcgttgg ctctgagaca ttgataataa atttttctca acagtgaaaa aaaaaaaaaa 2520
gaaaaaaaaa aaaaaa 2536

<210> 38
<211> 1350
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1747456CB1

<400> 38
cggaggagcc cgggcggggg ggaggaggag ggggaggagg gagcggagat ctcggggctc 60
ggagccggcc gccgctccgc tccgatcgct gtggggcttg gtttttgggg ggtgggggggg 120
cgggggggct cagatatgga ggcaaatggg agccaaggca cctcgggcag cgccaacgac 180
tcccagcacg accccggtaa aatgtttatc ggtggactga gctggcagac ctcaccagat 240
agccttagag actattttag caaatttgga gaaattagag aatgtatggt catgagagat 300
cccactacga aacgctccag aggcttcggt ttcgtcacgt tcgcagaccc agcaagtgta 360
gataaagtat taggtcagcc ccaccatgag ttagattcca agacgattga ccccaaagtt 420
gcatttcctc gtcgagcgca acccaagatg gtcacaagaa caaagaaaat atttgtaggc 480
gggttatctg cgaacacagt agtggaagat gtaaagcaat atttcgagca gtttggcaag 540

gtggaagatg caatgctgat gtttgataaa actaccaaca ggcacagagg gtttggcttt 600
gtcacttttg agaatgaaga tgttgtggag aaagtctgtg agattcattt ccatgaaatc 660
aataataaaa tggtagaatg taagaaagct cagccgaaag aagtcatgtt cccacctggg 720
acaagaggcc gggcccgggg actgccttac accatggacg cgttcatgct tggcatgggg 780
atgctgggat atcccaactt cgtggcgacc tatggccgtg gctaccccgg atttgctcca 840
agctatggct atcagttccc aggcttccca gcagcggctt atggaccagt ggcagcagcg 900
gcggtggcgg cagcaagagg atcaggctcc aacccggcgc ggcccggagg cttcccgggg 960
gccaacagcc caggacctgt cgccgatctc tacggccctg ccagccagga ctccggagtg 1020
gggaattaca taagtgcggc cagcccacag ccgggctcgg gcttcggcca cggcatagct 1080
ggacctttga ttgcaacggc ctttacaaat ggataccatt gagcaggtgc tttcgttgcc 1140
atctcactct gagagcatac ctggatgtcc aggcaagact gggcgaagtt tctgagtggc 1200
cctttgttta ggtgatgtcc tcagacctgg acccccacca gcctcactcc ccatcccaac 1260
cagagatggc tcacttcgga tcgagggttg actacatctc atcatctcac gaatctgctg 1320
taatataaga caacagcttt taaatgtgta 1350

<210> 39
<211> 2190
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1748626CB1

<400> 39
cctgcgggag ccccgtgccc gtcacgccgc cgcagccttc gccctggatg ccgctgctgc 60
cgccgccgtg ggactgtccc gggagcgggc cctggactac tacgggctgt acgacgaccg 120
tgggcgcccc tatggctacc cagctgtgtg tgaggaggac ctgatgcccg aggatgacca 180
gcgggccacg cgcaacctct tcattggtaa cctggaccac agcgtatctg aggtggagct 240
gcgaagggcc ttcgagaaat atggcatcat cgaggaggtg gtcatcaaga ggcctgcccg 300
tggccagggc ggtgcctatg ccttcctcaa gttccagaac ctggacatgg cccatagggc 360
taaggtggcc atgtcgggcc gagtgattgg tcgcaacccc attaagatag gctatggcaa 420
ggccaacccc accactcgtc tctgggtggg tggcctggga cctaacacgt cactggcggc 480
tctggcccga gagtttgacc gctttgggag cattcggacc attgatcacg tcaaaggaga 540
tagctttgcc tatattcagt acgagagctt ggacgcagcc caggccgcct gtgctaaaat 600
gaggggtttt cccttgggtg gaccagaccg caggctccgc gtggattttg ccaaagcaga 660
ggagactcgg taccccccagc agtaccagcc ctcgccactc cctgtgcatt atgagctgct 720
cacagatgga tacacccggc accgcaacct ggacgccgac ctggtgcggg acaggacgcc 780
cccacacctt ctgtactcag accgagaccg gacttttttg gaaggggact ggaccagccc 840
cagtaaaagc tctgaccgcc gaaacagcct tgagggctac agtcgctcag tgcgcagccg 900
gagtggtgag cgttgggggg cagatggaga ccgtggtttg cccaagccct gggaagagag 960
gcggaaacgg agaagccttt ccagtgaccg tgggaggaca acccattcac catatgagga 1020
acggagtagg accaagggca gtgggcagca gtcagagcgg ggctccgacc gcacccctga 1080
gcgcagccgc aaggagaacc actccagtga agggaccaag gagtccagca gcaactccct 1140
cagcaacagc agacatgggg ctgaggaacg gggccaccac caccaccacc acgaggctgc 1200
agactcttcc cacgggaaga aggcaagaga cagcgagcgc aatcaccgga ccacagaggc 1260
cgagcccaag cctctggaag agccaaaaca cgagaccaaa aagctgaaga atctttcaga 1320
gtacgctcag acactacagc tgggttggaa tgggcttctg gtgttgaaaa acagctgctt 1380
ccccacgtct atgcatatcc tagaggggga ccaggggtg atcagcagtc tcctcaaaga 1440
ccacacttct gggagcaagc tgacccagct gaagatcgcc cagcgccttc gactggacca 1500
gcccaagctt gacgaggtca cacgacgcat caagcagggg agccccaacg gctatgcggt 1560
cctcttagcc acccaggcaa cccccagtgg gcttggcact gaggggatgc ccacagtaga 1620
gcccggtctg cagaggcggc ttctcaggaa cctggtctcc tacttgaaac agaagcaggc 1680
cgcaggggtg atcagcttgc cagtgggggg gtccaagggc agagacggca caggcatgct 1740
ctacgccttc ccaccctgcg acttttccca gcagtacctc cagtcagcac taaggacatt 1800
gggcaagcta gaagaagaac acatggtgat agtcatcgtc agagacactg cctagcccaa 1860

gcctgtcttt cccagcgtca tgtttgtgtc acaaaagcag ttattttaaa atctgatccc 1920
ctctctaccc taccactttg gtttgaatta tctcctgggt tattttggtt catttgggtg 1980
gggatcaaag tcctgtccac caccaaaact aagttcttag attttgggggg attttttttt 2040
ttaaacgatg agaagggaat ccggttatgt tgatttctag tgtacaagat actgtctgct 2100
gtggttctgt attttttat tttttgacca actgtatgga aagttgtcag taaaaccttt 2160
gacagaggat ggatttttaa aaaaaaaaaa 2190

<210> 40
<211> 680
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 1879135CB1

<400> 40
gtctttggct gtgtggtctt aaatgtgttt ctaatgtgtg tgtcaaataa ttacctgtta 60
aacagactgc caatctggct gaagccaatg cttctgaaga agataaaatt aaagcaatga 120
tgtcgcaatc tggccatgaa tacgacccaa tcaattacat gaagaaacct ctaggtccac 180
cacctccatc ttacacgtgt ttccgttgtg gtaaacctgg acattatatt aagaattgcc 240
caacaaatgg ggataaaaac tttgaatctg gtcctaggat taaaaagagc actggaattc 300
ccagaagttt catgatggaa gtgaaagatc ctaatatgaa aggtgcaatg cttaccaaca 360
ctggaaaata tgcaatacca actatagatg cagaagcata tgcaattggg aagaaagaga 420
aacctccctt cttaccagag gagccatctt cttcctcaga agaagatgat cctatcccag 480
atgaattgtt gtgtctcatc tgcaaggata ttatgactga tgctgttgtg attccctgct 540
gtggaaacag ttactgtgat gaatgtaaga agtgctgaat cttggaagat gtatatttta 600
gaatatttgt atttacttgg aatggctctt cccaacctca tatgttttaa taataaaata 660
aataatgttg aaaaaaaaaa 680

<210> 41
<211> 1150
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2073417CB1

<400> 41
gacggaagtg cggtgttgag cgccggcggc tcgcgcccac gctgggccgg gagtcgaaat 60
gcttcccggt gccgggagtg agcgatgagc tggcttctgt tcctggccca cagagtcgcc 120
ttggccgcct tgccctgccg ccgcggctct cgcgggttcg ggatgttcta tgccgtgagg 180
aggggccgca agaccggggt ctttctgacc tggaatgagt gcagagcaca ggtggaccgg 240
tttcctgctg ccagatttaa gaagtttgcc acagaggatg aggcctgggc ctttgtcagg 300
aaatctgcaa gcccggaagt ttcagaaggg catgaaaatc aacatggaca agaatcggag 360
gcgaaagcca gcaagcgact ccgtgagcca ctggatggag atggacatga aagcgcagag 420
ccgtatgcaa agcacatgaa gccgagcatg gagccggcgc ctccagttag cagagacacg 480
ttttcctaca tgggagactt cgtcgtcgtc tacactgatg gctgctgctc cagtaatggg 540
cgtagaaggc cgcgagcagg aatcggcgtt tactgggggc caggccatcc tttaaatgta 600
ggcattagac ttcctgggcg gcagacaaac caaagagcgg aaattcatgc agcctgcaaa 660
gccattgaac aagcaaagac tcaaaacatc aataaactgg ttctgtatac agacagtatg 720
tttacgataa atggtataac taactgggtt caaggttgga agaaaaatgg gtggaagaca 780
agtgcaggga aagaggtgat caacaaagag gactttgtgg cactggagag gcttacccag 840
gggatggaca ttcagtggat gcatgttcct ggtcattcgg gatttatagg caatgaagaa 900

gctgacagat tagccagaga aggagctaaa caatcggaag actgagccat gtgactttag 960
tccttgggag aacttgagcc agcggctgtc ttgctgcctg tacttactgg tgtggaaaat 1020
agcctgcagg taggaccatt gcagtgatgg gcagatgcgt ctttcacacg gaatcaggca 1080
cagtggcctt ctgtgacatg tgtttataaa aaatggttaa gtatataata aattgaacat 1140
ctttgagatt 1150

<210> 42
<211> 2545
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2129080CB1

<400> 42
ggcagggagc ggagacggag gaggaggagg gagaggctga atgttggctc gggagacgta 60
cgaggaggac cgggagtacg agagccaggc caagcgtctc aagaccgagg agggggagat 120
cgactactcg gccgaggaag gcgagaaccg ccgggaagcg acgccccggg gcgggggcga 180
tggcggcggc ggcggccgga gcttctctca gccggaggca ggtggaagtc atcataaagt 240
ttctgtttca cccgtcgtcc atgttcgagg actctgtgaa tctgtggtgg aagcagacct 300
cgtggaagcg ctggaaaaat ttgggacaat atgctatgtg atgatgatgc catttaaacg 360
acaggctcta gtggaatttg aaaacataga tagtgccaaa gaatgtgtga catttgctgc 420
agatgaaccc gtgtacattg ctggtcaaca ggctttttc aactattcta caagcaaaag 480
gatcactcgg ccaggaaata ctgatgatcc atcaggaggc aacaaagttc ttctgctctc 540
aattcagaat ccgctttatc caattacagt ggatgtttta tatactgtat gcaaccctgt 600
tggcaaagtg caacgtattg ttatattcaa gagaaatggg atacaagcaa tggttgagtt 660
tgaatcagtc ctttgtgccc agaaagctaa agcagcactc aatggagctg atatatatgc 720
tggatgttgc acactaaaaa ttgaatatgc acggccaact cgtctaaatg ttattaggaa 780
tgacaatgac agttgggact acactaaacc atatttggga agacgagata gaggaaaggg 840
tcgccagaga caagccattt tgggagaaca cccttcttcg tttagacatg atggctatgg 900
atcccatggt ccattattgc ctttaccaag tcgttacaga atgggctctc gagatacacc 960
tgaacttgtt gcttatccat taccacaggc ttcttcctct tacatgcatg gaggaaatcc 1020
ctctggttca gttgtaatgg ttagtggatt acatcaacta aaaatgaatt gttcaagagt 1080
cttcaacctg ttctgcttat atggaaatat tgagaaggta aaatttatga agaccattcc 1140
tggtacagca ctggtagaaa tgggtgatga gtatgctgta gaaagagctg tcacacacct 1200
taataatgtc aaattatttg ggaaaagact taatgtttgc gtgtctaaac aacattcagt 1260
tgttccaagt caaatatttg agctggagga tggtaccagc agctacaaag attttgcaat 1320
gagcaaaaat aatcgcttta caagtgctgg ccaagcatct aagaatataa tccagccacc 1380
ctcctgtgtt ttgcattatt ataatgttcc attgtgtgtc acagaagaga ccttcacaaa 1440
gttgtgtaat gaccatgaag ttcttacatt catcaaatat aaagtgtttg atgcaaaacc 1500
ttcagccaaa acactttctg ggctattaga atgggagtgc aaaactgatg cagtagaagc 1560
ccttacggca ctgaatcact atcagataag agtgccgaat ggttccaatc cctatacatt 1620
gaagctttgc ttttctacat catcccattt ataagaagag aagagcatgt tagaatttat 1680
gttcaccttt attacaattt caaagctaca cttcattaaa aaaaaatcta aaatggttga 1740
tctcatgttg ccttgcttac tttaagatcc tgttctgtaa taaacatatt ttgccttgag 1800
taaatttgtt gtaagcttaa atattgaatt gttttcattt taagatagaa tatcataatg 1860
tagactatct acagcttcat tgtagattat acagatatat gatttctaac cttattactg 1920
gaatttttct tccacagtaa aaatatattt gcattcttaa tgctaattat ctgcaagtat 1980
tttttcattg tgtaagagat taatgcaggt gaaagtattg cattttaata tagaattcct 2040
attatatgtt tagatgttta agtatgttgc agttactcat attaaacata acttgtatat 2100
ttattatttt aatgaagttt gagaataacg ttacatatgt tgaattttaa gtactacaga 2160
tttaactgat tttatatttc tgaaaggcta acagacatgg atacacgtgt acagtatgca 2220
ttcaaactta tttaaattgg tgtatttttt tttaagtcac tgtccatttg tattgacatg 2280
cctctgtttc tagtccagtt tggagatttt ataaagttat aacaatgagt taatgtgttc 2340
attttcattt gttgcatgtg acttaaatac agctagtatt tggcattgag attttaatag 2400

aggttataat taacagttcc tcttatagat aataatctgg gatccatggg tgggcttcag 2460
agaatctgtg tacccctaa aattgtatgt agaattttgg atatttacat ttttatttt 2520
taactcttgg atccatcagt aatat 2545

<210> 43
<211> 907
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2472867CB1

<400> 43
caagcttggc gtttgtttgg tggggtcaca cgcgggttca acatgcgtat cgaaaagtgt 60
tatttctgtt cggggcccat ctatcctgga cacggcatga tgttcgtccg caacgattgc 120
aaggtgttca gattttgcaa atctaaatgt cataaaaact ttaaaaagaa gcgcaatcct 180
cgcaaagtta ggtggaccaa agcattccgg aaagcagctg gtaaagagct tacagtggat 240
aattcatttg aatttgaaaa acgtagaaat gaacctatca aataccagcg agagctatgg 300
aataaaacta ttgatgcgat gaagagagtt gaagaaatca aacagaagcg ccaagctaaa 360
tttataatga acagattgaa gaaaaataaa gagctacaga aagttcagga tatcaaagaa 420
gtcaagcaaa acatccatct tatccgagcc cctcttgcag gcaaagggaa acagttggaa 480
gagaaaatgg tacagcagtt acaagaggat gtggacatgg aagatgctcc ttaaaaatct 540
ctgtaaccat ttcttttatg tacatttgaa aatgcccttt ggatacttgg aactgctaaa 600
ttatttttatt ttttacataa ggtcacttaa atgaaaagcg attaaaagac atctttcctg 660
cattgccatc tacataatat cagatattac ggatgttaga ttgcatctca gtgttaaatc 720
tttactgata gatgtactta agtaaatcat gaaaattcta cttgtaacta tagaagtgaa 780
ttgtggacgt aaaatggttg tgctatttgg ataatggcac taggcagcat ttgtatagta 840
actaatggca aaaattcatg gctagtgatg tataaaataa aatattcttt gcagtaaaaa 900
aaaaaaa 907

<210> 44
<211> 1104
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2764755CB1

<400> 44
cttcgtttag gtcggctgga aattatgtcc tccgtcggtt ttccgcagtt tttccaccaa 60
gcgagatatt tttgggagtt attccctaaa taactgcatt atatgctcct ttcatgacga 120
aattgctgcc gtggagaaga ctggaggaaa ctcgaggaag agggagaagc cgacaagtgc 180
tcgacgggct aggaactgtc ctgcttgggt gttagcgttt cccgccgggc cagtaaggct 240
gagtgacccg gcgtggctac taggagaagg acgtacggtc ctgctagtag aggaatatgt 300
cgagtttctc tagggcgccc cagcaatggg ccacttttgc tagaatatgg tatctcttag 360
atgggaaaat gcagccacct ggcaaacttg ctgctatggc atctataaga cttcagggat 420
tacataaacc tgtgtaccat gcactgagtg actgtgggga tcatgttgtt ataatgaaca 480
caagacacat tgcattttct ggaaacaaat gggaacaaaa agtatactct tcgcatactg 540
gctacccagg tggatttaga caagtaacag ctgctcagct tcacctgagg gatccagtgg 600
caattgtaaa actagctatt tatggcatgc tgccaaaaaa ccttcacaga agaacaatga 660
tggaaaggtt gcatcttttt ccagatgagt atattccaga agatattctt aagaatttag 720
tagaggagct tcctcaacca cgaaaaatac ctaaacgtct agatgagtac acacaagaag 780
aaatagacgc cttcccaaga ttgtggactc cacctgaaga ttatcggcta taagagaata 840

agaattgcag aaaataacag tgaagtgatt gaaactttct tctgatgagt ttctctaacc 900
tacaggatgg agtaaaacaa ctgctacagt tcagcacctg ttttatgtgc cgaatcactg 960
tggggaaagg tcaggaaggt gtagtccttc aataggaaat tgtaattaaa atataatttt 1020
atagaaccat ttttatgtaa tctgatttga atgttatagt tgataataat aaaatcactt 1080
acttggttga ctaaaaaaaa aaaa 1104

<210> 45
<211> 910
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 2875939CB1

<400> 45
cccacgcgtc cgcgattcat agctcgcagg gtacgggcgc gcgtgcgcac tccgcagccc 60
gttcaggacc ccggcgcggg cagggcgccc acgagctggc tggctgcttg cacccacatc 120
cttctttctc tgggacctgg ggtcgcggtt acttgggctg gccggcgaac ccttgagtgg 180
cctggcgggg agcgggcctc gcgcgcctgg agggccctgt ggaacgaaga gaggcacaca 240
gcatggcaga aaaccgagag ccccgcggtg ctgtggaggc tgaactggat ccagtggaat 300
acacccttag gaaaaggctt cccagccgcc tgccccggag acccaatgac atttatgtca 360
acatgaagac ggactttaag gcccagctgg cccgctgcca gaagctgctg gacggagggg 420
cccggggtca gaacgcgtgc tctgagatct acattcacgg cttgggcctg gccatcaacc 480
gcgccatcaa catcgcgctg cagctgcagg cgggcagctt cgggtccttg caggtggctg 540
ccaatacctc caccgtggag cttgttgatg agctggagcc agagaccgac acacgggagc 600
cactgactcg gatccgcaac aactcagcca tccacatccg agtcttcagg gtcacaccca 660
agtaattgaa aagacactcc tccacttatc ccctccgtga tatggctctt cgcatgctga 720
gtactggacc tcggaccaga gccatgtaag aaaaggcctg ttccctggaa gcccaaagga 780
ctctgcattg agggtggggg taattgtctc ttggtgggcc cagttagtgg gccttcctga 840
gtgtgtgtat gcggtctgta actattgcca tataaataaa aaatcctgtt gcactagtaa 900
aaaaaaaaaa 910

<210> 46
<211> 733
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 3591363CB1

<400> 46
gttcgcgtcg tttccgtttc cggccgaggc tgcgggaaga tggcggcggc catggcagca 60
tcttccctga cggtcacctt agggcggctg gcgtccgcgt gcagccacag catcctgaga 120
ccttcggggc ccggagcagc ctccctttgg tctgcttctc gaaggttcaa ttcacagagc 180
acttcatatc taccaggata tgttcctaaa acatccctga gttcaccacc ttggccagaa 240
gttgttctgc cagacccagt tgaggagacc agacaccatg cagaggtcgt gaagaaggtg 300
aatgagatga tcgtcacggg gcagtatggc aggctctttg ccgtggtgca ctttgccagc 360
cgccagtgga aggtgacctc tgaagacctg atcttaattg gaaatgaact agaccttgcg 420
tgtggagaga gaattcgact ggagaaggtc ctgctggttg gggcagacaa cttcacgctg 480
cttggcaagc cactcctcgg aaaggatctt gttcgagtag aagccacagt cattgaaaag 540
acagaatcat ggccaagaat cattatgaga ttcaggaaaa ggaaaaactt caagaagaaa 600
agaatcgtca cgaccccgca gactgtcctc cggataaaca gcattgagat tgctccgtgt 660
ttgttgtgat taccgagtta atacttacaa aaggataaaa ataaactcct gcttcccaag 720

gaaaaaaaaa aaa 733

<210> 47
<211> 918
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 3702292CB1

<400> 47
cgcatgcggc cgagtgcggg actggggctc ctggctgtgg gtgtggtacc gaggcttcag 60
cgggtgccgc ccgcctagag ggagtggagc ggtgagcacg tcaggggtgg ggggcgcagg 120
tcaagctttc accagttttt aattctttga tggggtaaat ttgagcaatt ttctcgactt 180
gtcgacattc gttattaact gagcaggaat caggagagga acccggtcct ctccacacag 240
cccagcagag agcctacgac tagatttgca tctttacgtc ctgcgcggag gctgctacac 300
acatgcagaa gtcatgctgg tggcctggac agtgaaggga gagaagtgga tttgggagac 360
atttaggaga tggcaccaaa agcgaaggaa gctcctgctc atcctaaagc cgaagccaaa 420
gcgaaggctt taaaggccaa gaaggcagtg ttgaaaggtg tccgcagcca cacgcaaaaa 480
cagaagatcc gcatgtcact caccttcagg cggcccaaga cactgcgact ccggaggcag 540
cccagatatc ctcggaagag cacccccagg agaaacaagc ttggccacta tgctatcatc 600
aagtttccgc tggccactga gtcggccgtg aagaagatag aagaaaacaa cacgcttgtg 660
ttcactgtgg atgttaaagc caacaagcac cagatcagac aggctgtgaa gaagctctat 720
gacagtgatg tggccaaggt caccaccctg atttgtcctg ataaggagaa caaggcatat 780
gttcgacttg ctcctgatta tgatgctttc gatgttgtaa caaaattggg atcacctaaa 840
ctgagtccag ctggctaact ctaaatatat gtgtatcttt tcagcataaa aaaataatgt 900
ttttcataaa aaaaaaaa 918

<210> 48
<211> 2680
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 3778908CB1

<400> 48
agtggaactg gatcgggttt gctgccagcg gcgtgagctt cggccgccat tttacaacag 60
ctccactcgc gccggacaca gggagcagcg agcacgcgtt tcccgcaacc cgataccatc 120
ggacaggatt tctccgcctc agcccaacgg ggagatctct ggaaacatgg ctacagaaca 180
tgttaatgga aatggtactg aagagcccat ggatactact tctgcagtta tccattcaga 240
aaattttcag acattgcttg atgctggttt accacagaaa gttgctgaaa aactagatga 300
aatttacgtt gcagggctag ttgcacatag tgatttagat gaaagagcta ttgaagcttt 360
aaaagaattc aatgaagacg gtgcattggc agttcttcaa cagtttaaag acagtgatct 420
ctctcatgtt cagaacaaaa gtgccttttt atgtggagtc atgaagactt acaggcagag 480
agaaaaacaa gggaccaaag tagcagattc tagtaaagga ccagatgagg caaaaattaa 540
ggcactcttg gaaagaacag gctacacact tgatgtgacc actggacaga ggaagtatgg 600
aggaccacct ccagattccg tttattcagg tcagcagcct tctgttggca ctgagatatt 660
tgtgggaaag atcccaagag atctatttga ggatgaactt gttccattat ttgagaaagc 720
tggacctata tgggatcttc gtctaatgat ggatccactc actggtctca atagaggtta 780
tgcgtttgtc acttttgta caaaagaagc agctcaggag gctgttaaac tgtataataa 840
tcatgaaatt cgttctggaa aacatattgg tgtctgcatc tcagttgcca acaataggct 900
ttttgtgggc tctattccta agagtaaaac caaggaacag attcttgaag aatttagcaa 960

agtaacagag ggtcttacag acgtcatttt ataccaccaa ccggatgaca agaaaaaaaa 1020
cagaggcttt tgctttcttg aatatgaaga tcacaaaaca gctgcccagg caaggcgtag 1080
gttaatgagt ggtaaagtca aggtctgggg gaatgttgga actgttgaat gggctgatcc 1140
tatagaagat cctgatcctg aggttatggc aaaggtaaaa gtgctgtttg tacgcaacct 1200
tgccaatact gtaacagaag agattttaga aaaggcattt agtcagtttg ggaaactgga 1260
acgagtgaag aagttaaaag attatgcgtt cattcatttt gatgagcgag atggtgctgt 1320
caaggctatg gaagaaatga atggcaaaga cttggaggga gaaaatattg aaattgtttt 1380
tgccaagcca ccagatcaga aaaggaaaga aagaaaagct cagaggcaag cagcaaaaaa 1440
tcaaatgtat gacgattact actattatgg tccacctcat atgccccctc caacaagagg 1500
tcgagggcgt ggaggtagag gtggttatgg atatcctcca gattattatg gatatgaaga 1560
ttattatgat tattatggtt atgattacca taactatcgt ggtggatatg aagatccata 1620
ctatggttat gaagattttc aagttggagc tagaggaagg ggtggtagag gagcaagggg 1680
tgctgctcca tccagaggtc gtggggctgc tcctccccgc ggtagagccg gttattcaca 1740
gagaggaggt cctggatcag caagaggcgt tcgaggtgcg agaggaggtg cccaacaaca 1800
aagaggccgc ggggtacgtg gtgcgagggg tggccgcggt ggaaatgtag gaggaaagcg 1860
caaagctgat gggtacaacc agccagattc caagcggcgc cagaccaata atcagaactg 1920
gggctcccaa cccattgctc agcaaccgct ccaaggtggt gatcattctg gtaactatgg 1980
ttacaaatct gaaaaccagg agttttatca ggatactttt gggcaacagt ggaagtagaa 2040
acagtagggc ctctgtaaaa ttggagactg ataggttgat cagaaactca ccctaaatct 2100
gaacgggtgc cgctataatt tgtgacatct ggcaagattt ccctttatgt atatatttta 2160
acaatccgct tggacacgaa caaagccaca cttctaactg cttctggcga actgatttta 2220
tttttaattt ttttcaataa agatattctt agatactgaa agaaatagtt aatgagtttg 2280
catttgtgct tgagaaaatt tggctcaagt ccatttggct gtagtgtcaa cgatgtttcc 2340
agtagtgttt agatttggtg tcttcaaagg tagttgatta aaaccaagtg tgtctttaat 2400
atcttgtatc agaataactt tgtatgttac caacttaaat tgctagaata aggtaaattg 2460
atacacaact gctattttta atttagaact ttgacctaat ttgggttttc aaaaccattt 2520
tggctacttg tattctttat gctgttgttt atttcaataa aaaattcaca cctaaatgta 2580
tacttactaa aattgtgttt acaattcgtt tttcacaaaa tttcctgcaa atttggttca 2640
aattgtatag catgtcaagg ccaattaaag ggttttgtga 2680

<210> 49
<211> 2568
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 4163642CB1

<400> 49
ccaggctgcg ccagacagtg tagaacctgc ggcctcgatg tccttctccc gtgccctatt 60
gtgggctcgg ctcccggcgg ggcgccaggc tggccaccgg gcagccatct gctctgccct 120
tcgtccccac tttgggccct ttcccggggt tctggggcaa gtttctgtcc ttgccaccgc 180
ctcctcctct gcctccggtg gctccaaaat accaaacacg tccttgttcg tgcccctgac 240
tgtgaaacct cagggcccca gcgccgacgg cgacgtcggg gccgagctaa cccggcctct 300
ggacaagaat gaagtaaaga aggtcttaga caaattttac aagaggaaag aaattcagaa 360
actgggtgct gattatggac ttgatgctcg tctcttccac caagctttca taagctttag 420
aaattatatt atgcagtctc attccctgga tgtggacatt cacattgttt tgaatgatat 480
ttgcttcggt gcagctcatg cggatgattt attcccattt ttcttgagac atgccaaaca 540
aatatttcct gtgttggact gtaaggatga tctacgtaaa atcagtgact taagaatacc 600
acctaactgg tacccagatg ctagagccat gcagcggaag ataatatttc attcaggccc 660
cacaaacagt ggaaagactt atcacgcaat ccagaaatac ttctcagcaa agtctggagt 720
gtattgtggc cctctaaaat tactggcaca tgagatcttc gaaaagagta atgctgctgg 780
tgtgccatgt gacttggtga caggtgaaga gcgtgtgaca gttcagccaa atgggaaaca 840
ggcttcacat gtttcttgta cagttgagat gtgcagtgtt acaactcctt atgaagtggc 900
tgtaattgat gaaattcaaa tgattagaga tccagccaga ggatgggcct ggaccagagc 960

acttctagga ctgtgtgctg aagaggttca tttgtgtgga gaacctgctg ctattgacct 1020
ggtgatggag cttatgtaca caacggggga ggaagtggag gttcgagact ataagaggct 1080
tacccccatt tctgtgctgg accatgcact agaatcttta gataaccttc ggcctgggga 1140
ctgcattgtc tgttttagca agaatgatat ttattctgtg agtcggcaga ttgaaattcg 1200
gggattagaa tcagctgtta tatatggcag tctcccacct gggaccaaac ttgctcaagc 1260
aaaaaagttt aatgatccca atgacccatg caaaatcttg gttgctacag atgcaattgg 1320
catgggactt aatttgagca taaggagaat tatttttac tcccttataa agcccagtat 1380
caatgaaaag ggagagagag aactagaacc aatcacaacc tctcaagccc tgcagattgc 1440
tggcagagct ggcagattca gctcacggtt taaagaagga gaggttacaa caatgaatca 1500
tgaagatctc agtttattaa aggaaatttt gaagaggcct gtggatccta taagggcagc 1560
tggtcttcat ccaactgctg agcagattga aatgtttgcc taccatctcc ctgatgcaac 1620
actgtccaat ctcattgata tttttgtaga cttttcacaa gttgatgggc agtattttgt 1680
ctgcaatatg gatgatttta aatttttctgc agagttgatc cagcatattc cactaagtct 1740
gcgagtgagg tatgtttttct gcacagctcc tatcaacaag aagcagcctt ttgtgtgttc 1800
ttcactgtta cagtttgcca ggcagtatag caggaatgag cccctgacct ttgcatggtt 1860
acgccgatac atcaaatggc ctttacttcc acctaagaat attaaagacc tcatggatct 1920
tgaagctgtc cacgatgtct tggatcttta cttgtggcta agctaccgat ttatggatat 1980
gtttccagat gccagcctta ttcgagatct ccagaaagaa ctagatggta ttatccaaga 2040
tggtgtgcac aatatcacta aattgattaa aatgtctgag acgcataagc tgttgaattt 2100
ggagggcttt ccatcaggga gccagtcacg attgtcagga accttaaaga gccaagctag 2160
aaggacacgc ggcaccaaag ctctagggag taaagctact gagccaccca gccccgatgc 2220
aggagagctg tcccttgctt ccagattggt gcagcaagga ctcctcactc cagacatgct 2280
gaaacagcta gaaaaagagt ggatgacaca acaaactgaa cacaacaaag aaaaaacaga 2340
gtctgggact catccaaaag ggacgagaag aaagaagaag gaacctgatt cggactagtt 2400
ttctgttcct gtttttttttt tttatttaat tttgcaaata aaaatttatt ttgaatcctt 2460
tttcctcata tgcatttact ccctcctcta gtattgtggc tatctggtac tgggggattt 2520
ttggtgtgtg tgtgtgtttg tgtgtgtgtt tttttgtttg ttttttcc 2568

<210> 50
<211> 847
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte Identification No.: 4906154CB1

<400> 50
caaatggcgg atgacgccgg tgcagcgggg gggcccgggg gccctggtgg ccctgggatg 60
gggccggagg aaggccgagg ataaggagtg gatgcccgtc accaagttgg gccgcttggt 120
caaggacatg aagatcaagt ccctggagga gatctatctc ttctccctgc ccattaagga 180
atcagagatc attgatttct tcctgggggc ctctctcaag gatgaggttt tgaagattat 240
gccagtgcag aagcagaccc gtgccggcca gcgcaccagg ttcaaggcat ttgttgctat 300
cggggactac aatggccacg tcggtctggg tgttaagtgc tccaaggagg tggccaccgc 360
catccgtggg gccatcatcc tggccaagct ctccatcgtc cccgtgcgca gaggctactg 420
ggggaacaag atcggcaagc cccacactgt cccttgcaag gtgacaggcc gctgcggctc 480
tgtgctggta cgcctcatcc ctgcacccag gggcactggc atcgtctccg cacctgtgcc 540
taagaagctg ctcatgatgg ctggtatcga tgactgctac acctcagccc ggggctgcac 600
tgccaccctg ggcaacttcg ccaaggccac ctttgatgcc atttctaaga cctacagcta 660
cctgaccccc gacctctgga aggagactgt attcaccaag tctccctatc aggagttcac 720
tgaccacctc gtcaagaccc acaccagagt ctccgtgcag cggactcagg ctccagctgt 780
ggctacaaca tagggttttt atacaagaaa aataaagtga attaagcgtg ttaaaaaaaa 840
aaaaaaa 847